

Table of Contents

Table of Contents	i
Preface from the Chair	ii
Acknowledgements	iii
Executive Summary	1
Establishment and Membership	4
Charge, Vision, and Process	4
Guiding Principles for Michigan's Educator Evaluation System	5
Definitions	6
Teacher Evaluation Overview	7
Teacher Evaluation: Observations of Practice	9
Teacher Evaluation: Student Growth	13
Combining and Implementing the Practice and Student Growth Components	22
Administrator Evaluation Recommendations	24
Professional Teaching Certificate	28
LEA Waiver Processes and Principles	28
Developing, Implementing, and Revising the Evaluation System	30
Appendix A: Council Members' Biographies	35
Appendix B: Annotated Agendas	36
Appendix C: In-Meeting Consultations	102
Appendix D: Out-of Meeting Consultations	107
Appendix E: Research and Resources	111
Appendix F: Administrator Evaluation Tools	117
Appendix G: ISR's Administrator Evaluation Tool Structured Review Report	118
Appendix H: Validity and Reliability Required for Waiver Approval	148
Appendix I: Possible Infrastructure Frameworks	149

Preface from the Chair

In a growing body of research, study after study has shown the power of skillful teaching to affect student learning. The details are dramatic: A child who receives high quality instruction three years in a row will have learned significantly more—comparable to having had many more months of instruction—than a child whose teachers for three years have lacked skill and capacity. Skillful teaching has a powerful impact on students' learning. Teachers who are prepared and supported to teach attentively are able to connect with students and make complex ideas and content learnable. They are able to form relationships with children and support them to master the academic and personal skills essential for success. Their professional skills enable them to care effectively for students' academic, social, emotional, and physical well being and development.

Every child in Michigan deserves skillful teachers, not just some of the time, but each and every year. And every teacher deserves the opportunity to develop and continue to refine his or her professional skill—to receive targeted feedback and professional learning opportunities to improve instruction. Our commitment to this goal animated the work of the Michigan Council for Educator Effectiveness (MCEE).

Over the last 18 months, my five MCEE colleagues and I developed the recommendations for Michigan's educator evaluation system contained in this report. We examined research. We consulted with other states. We commissioned a pilot study. We engaged practicing educators. We requested expert guidance. We read and questioned and sought advice.

Through it all, we were guided by a shared vision that we established early on:

The Michigan Council for Educator Effectiveness will develop a fair, transparent, and feasible evaluation system for teachers and school administrators. The system will be based on rigorous standards of professional practice and of measurement. The goals of this system are to contribute to enhanced instruction, improve student achievement, and support ongoing professional learning.

This vision demonstrates our deep commitment to educators, to their continued growth and improvement as professionals, and to the academic outcomes and future success of Michigan's children. We believe that a system of educator evaluation should be grounded in improvement-focused feedback, not in punitive action. The system should, at its core, serve to raise the performance of **all** educators because this is what our students need and deserve. Our young people are the future citizens of our state and our nation; the quality of their education is our most important investment.

Making skillful teaching the norm will not only enhance the life chances and fulfillment of Michigan's youth, but also boost our strength and capacity as a state.

Deborah Loewenberg Ball
Chair, Michigan Council for Educator Effectiveness
Dean, University of Michigan School of Education

Acknowledgements

The MCEE's work would not have been possible without the support of the following organizations and people:



The Advisory Council to the MCEE

- Tonya Allen
- Amber M. Arellano
- Ernst A. Bauer
- Richard S. Carsten
- William C. Chilman, IV
- Dan L. DeGrow (chair)
- Ingrid J. Guerra-Lopez
- John F. Haan
- Krista L. Hunsanger
- Mary A. Kovari
- Barbara F. Mays
- Kirstin G. Queen
- Colin Ripmaster
- Tammy M. Wagner
- Nathan R. Walker
- Matthew T. Wandrie

Ongoing consultants to the MCEE

- Brian Rowan, Burke A. Hinsdale Collegiate Professor, School of Education, and research professor, Institute for Social Research, University of Michigan
- Suzanne Wilson, University Distinguished Professor and chair of the Department of Teacher Education, Michigan State University

External reviewers of the MCEE's final recommendations

- Naomi Chudowsky, consultant, Caldera Research; former study director at the Board on Testing and Assessment, National Research Council
- Stuart W. Elliott, director, Board on Testing and Assessment, National Research Council
- Michael J. Feuer, dean at the Graduate School of Education and Human Development, George Washington University; former executive director of the Division of Behavioral and Social Sciences and Education, National Research Council
- David Stuit, founding partner, Basis Policy Research

MCEE project manager

- Corinne F. Mehan

Executive Summary

The Michigan Council for Educator Effectiveness developed the following key recommendations for each section of its charge.

TEACHER EVALUATION

Teachers play a critical role in student learning and skillful teaching has a significant impact on students' opportunities and growth. The evaluation of teachers' performance is to be composed of evidence about two key aspects of their work: (1) their practice and (2) their students' growth.

Key Recommendations for Observations of Practice

- ❖ Classroom teaching must be observed using one of the four piloted observation tools: Charlotte Danielson's Framework for Teaching, Marzano Teacher Evaluation Model, The Thoughtful Classroom, or 5 Dimensions of Teaching and Learning. If final results from the pilot study produce evidence that suggests that any of these tools is less reliable or practical, this information should be taken into account.
- ❖ One of these tools will be selected to be the state tool, based on a competitive RFP process. The state will provide sufficient base funding to support Local Education Agencies' (LEAs') use of the state-selected tool with full fidelity. The state will provide the technical support and training for the state-selected tool.
- ❖ LEAs may choose to use one of the other three piloted observation tools instead, but must pay for any expenses above the base funding supplied by the state for the state-selected tool.
- ❖ Any observer in an LEA must be trained in that LEA's observation protocol, whether the LEA is using the state tool or one of the other piloted tools. To support fidelity in the use of the observation tool, the training must be provided by the observation tool vendor. The training regime for an observer must be equivalent to that which is offered for the state-selected tool in terms of time and content.
- ❖ Multiple observations must be conducted across the school year. At least one observation must be unscheduled.
- ❖ Qualified peers may conduct some of the observations.
- ❖ The administrator responsible for a teacher's evaluation must conduct at least one of the observations and must use results of others' observations (if applicable) in making an overall appraisal of the teacher's practice.
- ❖ In addition to receiving the above-referenced vendor-provided training on the observation tool, the administrator responsible for a teacher's evaluation must receive adequate training (which includes opportunities to practice) in coaching and providing feedback.
- ❖ All teacher observation data collected with the state-selected observation tool shall be reported to the state for research purposes to support continuous improvement of the educator evaluation system.

Key Recommendations for Student Growth

- ❖ The state should continue to develop or select, as well as support, all aspects of assessments that are aligned to state-adopted content standards in all core content areas (i.e., English language arts, mathematics, science, and social studies).
- ❖ The state should also develop or select assessments aligned to state-adopted content standards in high-volume non-core content areas where state-adopted content standards exist (e.g., arts, health and physical education, career and technical education, and many high school electives).
- ❖ In addition, the state should provide guidelines for:
 - Evaluating the quality of third party or locally developed assessments.
 - Training on adequate rigor in the development and measurement of student learning objectives.
- ❖ The state should, in all cases that are possible and professionally responsible, produce value-added modeling (VAM) scores for educators on state-provided assessments in the core content areas.
- ❖ For teachers in core content areas in grades for which there are growth data available from state-mandated assessments (currently reading and mathematics in grades 4–8, but likely to change over time), at least half of the teachers' student growth component should be based on state-provided VAM scores.

- ❖ For teachers of other subject areas, LEAs should have the discretion to adopt state-provided VAM or growth data.
- ❖ State-provided VAM or growth data in core content areas may be used in a teacher's evaluation using information from that teacher's students, even if the teacher does not teach in one of the core content areas. This means that teachers may be evaluated, in part, for the learning of their own students, even in subject areas that they do not directly teach. This may be done as long as the teacher knows that he or she is expected to be contributing to students' growth in core content areas and there is a reasonable connection of the core content to the teacher's actual teaching assignment.
- ❖ In order to promote collective work on instruction within school buildings, school-level VAMs may be used for individual teachers' evaluations if there is a reasonable connection of the core content to the teacher's actual teaching assignment. This means that teachers may be evaluated, in part, for the learning of students whom they do not directly teach. However, school-level VAMs may not comprise more than 10% of the individual teacher's student growth component.
- ❖ In content area assignments for which there is no state-provided VAM or growth data available, teachers should be evaluated based on alternate measures of student growth that meet the guidelines for rigorous and appropriate assessment of student growth in the applicable subject area.

ADMINISTRATOR EVALUATION

Administrators play a central role in high quality instruction. They support teachers, provide feedback, and enable and enhance professional learning communities. They manage resources, work with parents, and play a critical role in communicating between the school and the community.

Key Recommendations for Administrator Evaluation

- ❖ LEAs must choose from one of two administrator evaluation tools: MASA's School ADvance Administrator Evaluation Instrument or Reeves Leadership Performance Rubric. One of these tools will be selected to be the state tool, based on a competitive RFP process. The state will provide sufficient base funding per administrator to support LEAs' use of the state-selected tool with full fidelity. The state will also provide the technical support and training for the state-selected tool.
- ❖ LEAs may choose to use the other tool instead. An LEA that chooses to use the other tool must pay for any expenses above the base funding supplied by the state for the state-selected tool.
- ❖ Administrators must be evaluated on both their practice and student growth measures.
- ❖ In addition to student growth measures, administrators must be evaluated by at least the following evidence for each school (or LEA, for superintendents or central office personnel): proficiency of their skill in evaluating teachers; progress made in the school improvement plan; attendance rates; and student, parent, and teacher feedback. LEAs can choose to include other information.
- ❖ The school board must evaluate superintendents.

CATEGORIZING AND PROVING FEEDBACK TO TEACHERS AND ADMINISTRATORS THROUGH ANNUAL EVALUATION SUMMARIES

Evidence about an educator's performance will be combined into a final evaluation rating using an approach that maps the evidence onto a rubric with two axes: (1) quality of practice (based on observations) and (2) quality of student outcomes (based on student growth data). The MCEE recommends using three categories for rating teachers and administrators:

Professional: A teacher or administrator rated **professional** has exhibited the knowledge and capabilities expected of a skillful educator. We expect that educators who are extraordinary—as well as educators who are competent—would fall into this category. We also assume that even highly skilled practice can be improved. Because the purpose of educator evaluation is improvement, and because all educational practice benefits from ongoing refinement, a rating of “professional” should always be accompanied by specific feedback for development. A teacher or administrator rated as professional for three straight years may pursue opportunities for advanced roles or leadership. In addition, an educator rated as professional for three straight years may be evaluated on an alternating year basis in subsequent years and receive two-year goals for enhancement.

Provisional: A teacher or administrator rated **provisional** has exhibited some professional knowledge and skill, but has specific substantial identified weaknesses that should be addressed through feedback and targeted professional development. We expect that this category might include two groups of educators: beginning teachers and administrators who have yet to develop the level of skill necessary for professional practice and experienced teachers and administrators who struggle with mastering professional knowledge and skills. This rating is intended to be a serious signal that the educator's practice requires significant improvement in specific areas and the rating should be accompanied by clear feedback about areas that need focused attention. A teacher or administrator rated as provisional or below for three straight years should be counseled out of his or her current role.

Ineffective: A teacher or administrator rated **ineffective** has exhibited performance that has specific critical identified weaknesses. The educator should be placed on urgent notice that significant improvement must be achieved in the specific areas of underperformance within two years. An educator who receives an ineffective rating for two years in a row should be terminated from further employment as a teacher (for teachers) or as an administrator (for administrators) in his or her current LEA.

Each educator will receive a summative rating that represents the overall assessment of his or her performance—both in terms of practice and student growth—over the past year. More important, however, is that every annual evaluation—even of skillful educators—should specify goals for the continued development of practice. These goals should be developed by the evaluator in consultation with the educator being evaluated. In addition, educators should be provided with specific suggestions, resources, and opportunities for the learning needed to achieve those goals.

PROFESSIONAL TEACHING CERTIFICATION

Currently, Michigan teachers initially receive certification for five years, with an opportunity to renew this first certification for three additional five-year terms if necessary. This initial license to teach is labeled “provisional certification.” The next level of certification is called “professional.”

Key Recommendations for Professional Teaching Certification

- ❖ Teachers wishing to move from a provisional to a professional certificate must receive a professional rating for three successive years immediately prior to applying for professional certification.
- ❖ Teachers who do not demonstrate three successive years of professional ratings may apply for the renewal of their provisional certification until they either:
 - Achieve three successive years of professional ratings immediately prior to applying for their professional certification; or
 - Have three non-successive years of professional ratings but have the recommendation of their current principal.

WAIVER PROCESS

Although the MCEE is committed to creating a quality system of educator evaluation, the council recognizes that some LEAs have used time and resources to develop their own evaluation systems. LEAs that are able to demonstrate the effectiveness of their local evaluation tools and policies may apply for a waiver from the state system.

Key Recommendations for the Waiver Process

- ❖ In order to receive a waiver from using the state recommended process, entities requesting a waiver must demonstrate that their processes and systems have the same level of quality and rigor as those adopted in LEAs following the state requirements.
- ❖ If an LEA submits an adapted form of a commercial evaluation system, the LEA must demonstrate how the adaptations do not threaten the validity of the inferences based on use of the instrument.
- ❖ If an LEA is using an evaluation system that does not have available documentation about its validity and reliability, the LEA must submit a plan for how it will gather relevant data on the system's technical soundness. LEAs will receive probationary approval for their waiver, contingent on providing evidence of their system's reliability and validity within three years.

Establishment and Membership

The Michigan Council for Educator Effectiveness (MCEE) was established in June 2011 as part of Michigan's teacher tenure reform efforts (PA 102 of 2011). Council members were appointed in September 2011, and the legislature appropriated funding in mid-December 2011. The MCEE was a temporary commission with a life of no more than two years and officially disbanded in June 2013.

The council had five voting members, three of whom were appointed by Governor Rick Snyder, and one each by Senate Majority Leader Randy Richardville and Speaker of the House Jase Bolger. Governor Snyder appointed Deborah Loewenberg Ball, dean of the University of Michigan School of Education, as chair of the MCEE. In addition to Ball, the governor appointed Mark Reckase from Michigan State University's College of Education and Nick Sheltrown from National Heritage Academies in Grand Rapids. Majority Leader Richardville appointed David Vensel, a principal from Jefferson High School in Monroe, and Speaker Bolger appointed Jennifer Hammond, a principal from Grand Blanc High School. Joseph Martineau served on the MCEE without a vote and was the designee of the Michigan Department of Education's superintendent of public instruction. (See Appendix A for biographies of council members.)

Charge, Vision, and Process

CHARGE

The MCEE was charged by law with an ambitious agenda, one that has tremendous significance for the educational opportunities and outcomes of our state's children. This report represents fulfillment of the charge to submit to the State Board of Education, the Governor, and the state legislature recommendations for all of the following:

- A state evaluation tool for teachers.
- A student growth and assessment tool.
- A state evaluation tool for school administrators.
- Changes to the requirements for a professional teaching certificate.
- A process for evaluating and approving local evaluation tools for teachers and administrators that are consistent with the state evaluation tool for teachers and administrators and the act.

VISION

From its inception, the MCEE focused its work on developing a system of evaluation for Michigan's educators that would be focused on improving instruction and advancing student learning—not on punitive action. As such, the following common vision grounded the MCEE's efforts:

The Michigan Council for Educator Effectiveness will develop a fair, transparent, and feasible evaluation system for teachers and school administrators. The system will be based on rigorous standards of professional practice and of measurement. The goals of this system are to contribute to enhanced instruction, improve student achievement, and support ongoing professional learning.

PROCESS

The council completed its work through two primary processes: (1) conducting a pilot study of educator effectiveness tools to provide data on implementation and validity and crucial feedback from education professionals; and (2) meetings, consultation, and research.

Pilot Study: The 2012–2013 Pilot of Educator Effectiveness Tools, which took place from September 2012 to June 2013 in thirteen Local Education Agencies (LEAs) across Michigan, informed these recommendations and will support the implementation of a fair, reliable, and effective approach to

educator evaluation in Michigan. Throughout the year, the pilot study team regularly provided the MCEE with updates and insights from their work. For example, learning about the kinds of training that different vendors provided helped the council to recognize the importance of careful training in order for any tool to be used effectively. The MCEE also learned about the challenges of managing the numbers of observations, and of entering data from the observations. Because the pilot study team was in regular contact with the field, they were also able to keep the council apprised of concerns raised by teachers and administrators alike. Although this report is informed by the pilot study, it does not represent a complete report on the pilot study. A final report will be submitted separately to the council by the Institute for Social Research (ISR) at the University of Michigan, which designed, implemented, and managed the pilot study.

Meetings, Consultation, and Research: The MCEE met regularly (at least twice a month for 3 to 4 hours) in whole group meetings. (Appendix B summarizes each council meeting's presentations and discussions.) In addition, subgroups of the council worked on specific high priority tasks such as investigating how other states have tackled particular challenges, exploring alternative methods for tracking student growth, and researching approaches to administrator evaluation. These subgroups met several additional times between council meetings. In order to provide opportunities for stakeholder and expert input, meetings were regularly open to the public and the MCEE website (www.mcede.org) posted updates and information, and solicited comments and suggestions.

In addition to the work of its six members, the MCEE benefitted from the input of expert consultants, all of whom are national leaders in areas crucial to the council's work. These experienced scholars and practitioners provided valuable insight into education policies, reforms, and initiatives that are taking place in Michigan and in other states. Council members consulted with nearly 50 experts from more than 10 states (see Appendices C and D). They also referred to research and reports from a wide range of organizations and commissions around the country that have worked extensively to understand educator evaluation and to implement evaluation systems (see Appendix E). Research and consultants provided the MCEE with a wealth of knowledge regarding observation tools, student growth models, pilots, and both the successes and concerns of other states throughout similar processes.

Guiding Principles for Michigan's Educator Evaluation System

To ensure a strong and improvement-focused system, the MCEE developed the following design principles for Michigan's educator evaluations:

- The system should support clear and rigorous expectations for educational practice in teaching and administration.
- The resources—money, time, attention—required to implement the educator evaluation system should be carefully balanced so as to maximize the value of the system while preserving resources for other requirements. For the evaluation system to support improvement, resources—including money and time—must be preserved for professional development.
- The system should be designed to support efforts toward improvement rather than compliance.
- The system should develop local capacity to improve teaching and learning at all levels and in all LEAs.
- The system for evaluating educators' performance should be based on multiple measures. This is vital in order to take into account the different dimensions of skilled educational practice and to increase the validity and reliability of evaluation. However, the measures used should also be limited so as to ensure efficiency when possible.
- The system should be designed to improve teaching and learning in the state, and structured to support ongoing educator learning and development.

In addition, the MCEE agreed on five specific parameters necessary in order to ensure that the system contributes to improvement of instruction and student learning:

1. An educator evaluation system that can support educational improvement depends on a well-articulated and shared consensus about goals for students' learning and principles for assessing their growth toward those goals. A rigorous educator evaluation system, thus, is dependent on clearly articulated student standards.
2. Similarly, an educator evaluation system that can support educational improvement depends on the articulation of clear standards of professional practice.
3. To support improvement, educator evaluation must provide specific substantive feedback that is accompanied by targeted and specific professional learning opportunities. Ratings alone do not supply sufficient information for improvement.
4. Individual educators' evaluations must be treated as confidential personnel information. They should be used to focus professional development and guide improvement-focused actions, but they should not be made public. The goal is improvement, not embarrassment.
5. Educator evaluations must not be used to determine merit pay. There is insufficient evidence to support pay for performance programs at this time. Research in education and other fields suggests that performance-related pay and the monitoring systems that come with it can backfire, decreasing motivation and quality performance.

Definitions

The approach to educator evaluation uses terms—e.g., “teacher,” “administrator,” “reliability,” and “validity”—that are understood in a variety of ways. To provide clarity, the MCEE defined this set of key terms:

What is a teacher?

A teacher is an individual who is directly responsible for instruction aimed at helping a group of students reach goals defined in a well-specified curriculum over an extended period of time such as a quarter, semester, or academic year.

What is a teacher leader or master teacher?

A teacher leader is a teacher whose performance has been rated professional for at least three years in a row and who has additional skills in supporting the development and improvement of practice.

What is a school administrator?

A school administrator is an educator whose role includes major responsibility for the leadership and management of a school building or of a key instructional function of an LEA.

What is reliability?

Reliability refers to the degree to which a measurement of skill or knowledge is consistent and stable in yielding similar results under varying conditions. Most simply put, a measurement is reliable if it is consistent within itself, between observers/raters, and across time.

What is validity?

Validity refers to the degree to which a measurement of skill or knowledge actually measures what it claims to measure. Validity is also the extent to which inferences, conclusions, and decisions made on the basis of measurements are appropriate and meaningful.

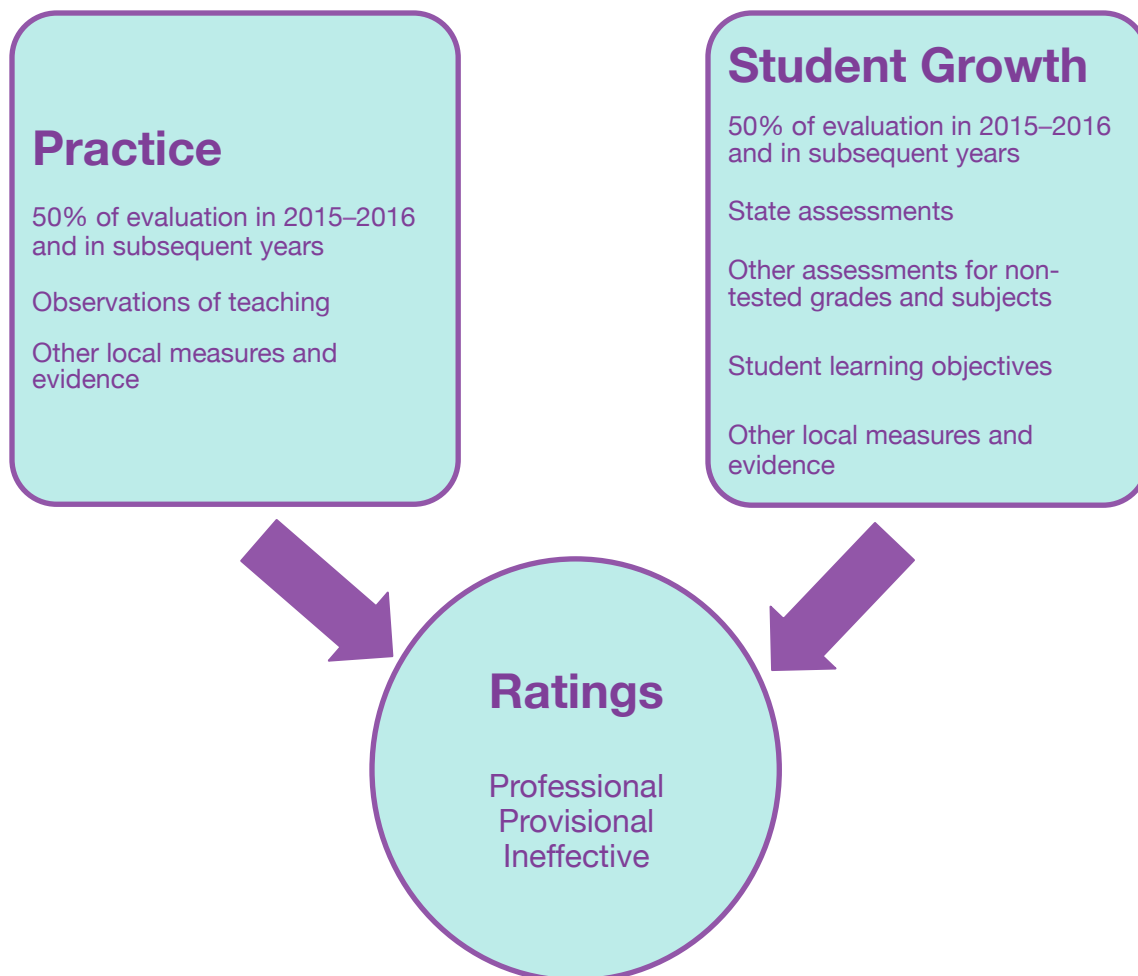
Teacher Evaluation Overview

FRAMEWORK FOR EVALUATING TEACHERS

Teachers play a critical role in student learning and skillful teaching has a significant impact on students' opportunities and growth. Every child in Michigan deserves skillful teachers each and every year. And every teacher deserves the opportunity to develop and continue to refine his or her professional skill—to receive targeted feedback and professional learning opportunities to improve instruction.

The evaluation of teachers' performance is to be composed of evidence about two key aspects of their work: (1) their practice and (2) their students' growth. "Practice" refers to the work that teachers do to prepare and conduct instruction and to assess, communicate about, and improve students' learning. This includes working with students, administrators, caregivers, and colleagues. "Student growth" represents the change in students' knowledge and skill across a school year.

The following figure outlines the components of the teacher evaluation system.



CATEGORIZING TEACHERS

Although the original legislation referred to a four-category system for reporting the results of teacher evaluation, the MCEE had the following significant concerns about the lack of reliability and usefulness of using four categories to rate teachers' performance:

- There is no empirical or theoretical basis for using four categories.
- The "highly effective" category suggests that these educators do not need to continue to improve. It might also lead attention to be focused on ratings (similar to students seeking A+ grades) and not on specific feedback for growth.
- The MCEE believes that it is not tenable to use a system of consequentially different ratings that cannot be fully justified. We urge instead a shift in emphasis to clear improvement plans for all teachers including those whose performance is judged as within the professional range.
- The MCEE seeks to support continuous improvement of practice. Because skillful teaching is complex, no teacher is "done" or finished learning. "Improvement" should be the norm instead of something seen as negative.
- Most urgently needed from ratings is a way to identify two groups: teachers whose performance is in need of immediate and significant change, and teachers whose performance is simply unacceptable.

DEFINING THE THREE CATEGORIES

The MCEE recommends the use of three evaluation ratings: professional, provisional, and ineffective. These ratings refer to evaluation outcomes, not certification levels. These three categories are considered as ordered qualitative categories with the following descriptions:

Professional: A teacher rated **professional** has exhibited the knowledge and capabilities expected of a skillful educator. We expect that teachers who are extraordinary—as well as teachers who are competent—would fall into this category. As such, the MCEE believes that the majority of teachers will receive a professional rating. We also assume that even highly skilled practice can be improved. Because the purpose of educator evaluation is improvement, and because all educational practice benefits from ongoing refinement, a rating of "professional" should always be accompanied by specific feedback for development. A teacher rated as professional for three straight years may pursue opportunities for advanced roles or leadership.¹ In addition, a teacher rated as professional for three straight years may be evaluated on an alternating year basis in subsequent years and receive two-year goals for enhancement.

Provisional: A teacher rated **provisional** has exhibited some professional knowledge and skill, but has specific substantial identified weaknesses that should be addressed through feedback and targeted professional development. We expect that this category might include two groups of teachers: beginning teachers who have yet to develop the level of skill necessary for professional practice and experienced teachers who struggle with mastering professional knowledge and skills. This rating is intended to be a serious signal that the educator's practice requires significant improvement and growth in specific areas and the rating should be accompanied by clear feedback about areas that need focused attention. A teacher rated as provisional or below for three straight years should be counseled out of his or her current role.

Ineffective: A teacher rated **ineffective** has exhibited performance that has specific critical identified weaknesses. The teacher should be placed on urgent notice that significant improvement must be achieved in the specific areas of underperformance within two years. A teacher who receives an ineffective rating for two years in a row should be terminated from further employment as a teacher in his or her current LEA.

¹ If LEAs wish to identify teachers whose performance is consistently exemplary in order to select master teachers or for some other purpose outside of the regular evaluation system, they may do so.

In the first two years of the implementation of the new educator evaluation system, the state should initiate a standard-setting process that will enable clearer specification of the qualities of practice associated with each of these levels. Specific standards for instructional practice will help to support the emphasis on improvement that will build a strong system of teacher and student learning.

Teacher Evaluation: Observations of Practice

OVERVIEW

The MCEE is committed to institutionalizing teacher observations as part of Michigan's educator evaluation system in a rigorous, professionally responsible, and legally defensible way. Because so many states have recently created such systems, the council gathered information from across the country about the components of such systems, the tools available, the measurement challenges associated with educator observations, the processes and resources needed to guarantee rigorous use of these measures, and the lessons that other states have learned along the way.

Many observation and evaluation systems are currently available. Some have been developed by researchers, others by professional developers, others by educators committed to providing sound support for early career teachers. Several states—Rhode Island, North Carolina, and Colorado, for example—have developed their own protocols (often adapting aspects of other widely used observation tools). Evidence about these tools' reliability and validity is not available.

After reviewing a number of possible observation protocols in use across the country, considering available evidence about their basis and use, examining them for both comprehensiveness and efficiency, and learning from LEAs that had already implemented some form of teacher evaluation, the MCEE chose the following four teacher observation tools to be piloted and studied in the LEAs participating in the pilot study:

- Charlotte Danielson's Framework for Teaching
- Marzano Teacher Evaluation Model
- The Thoughtful Classroom
- 5 Dimensions of Teaching and Learning

In July and August of 2012, ISR processed more than 70 applications from Michigan LEAs interested in participating in the pilot. ISR selected and recruited a final sample of thirteen LEAs based on a number of factors including school size, free and reduced lunch percentages, and geographic location.

The pilot LEAs included:

- Big Rapids Public Schools
- Cassopolis Public Schools
- Clare Public Schools
- Farmington Public Schools
- Garden City Public Schools
- Gibraltar School District
- Harper Creek Community Schools
- Leslie Public Schools
- Marshall Public Schools
- Montrose Community Schools
- Mt. Morris Consolidated Schools
- North Branch Area Schools
- Port Huron Area School District

Participating LEAs pledged to comply with the pilot design and protocols, agreeing—among many other requests—to adopt the assigned observation tool, participate in observation tool training, administer the pilot testing regime, and allow ISR researchers to conduct teacher observations, surveys, and interviews throughout their schools. The pilot design also required administrators to complete three observation sessions with each classroom teacher using their observation protocol. ISR team members also met regularly with participating LEAs to gather feedback and insight into LEA experience with the tools.

KEY RECOMMENDATIONS

Using data from the pilot study, research, consultations, and council meetings, the MCEE developed the following key recommendations regarding observations and other evidence that can be used to provide teachers with data about their practice.

Observation Tools, Training, and Support

- ❖ Current data from the pilot study suggests little significant difference among the tools. Therefore, the state should select one of the four piloted observation tools—Charlotte Danielson’s Framework for Teaching, Marzano Teacher Evaluation Model, The Thoughtful Classroom, or 5 Dimensions of Teaching and Learning—through a competitive RFP process. If final results from the pilot study produce evidence that suggests that any of these tools is less reliable or practical, this information should be taken into account.
- ❖ The state will provide sufficient base funding per teacher to support LEAs’ use of the state-selected tool with full fidelity.
- ❖ The state will provide technical support and training for the state-selected observation tool. This includes gathering and managing the observational data for LEAs that use that system.
- ❖ The state should provide hosting and data analysis services for the state-selected observation tool for all Michigan LEAs that adopt that tool. Such data should be used by the state for research purposes, including monitoring the reliability and validity of the observation tool.
- ❖ Any LEA that chooses to use one of the other three piloted tools must pay for any additional expenses above the base cost supplied by the state, including costs of technical support, training, data management, data analysis, and reporting.
- ❖ Observers must be trained in the use of their LEA’s observation tool, whether the LEA is using the state-selected tool or one of the other piloted tools. To support fidelity in the use of the observation tool, the training must be provided by the observation tool vendor. The training regime for an observer must be equivalent to that which is offered for the state-selected tool in terms of time and content. The goal of the training should be to ensure a baseline level of inter-rater agreement/reliability (i.e., after completing training, two observers should be able to observe the same lesson by the same teacher and their scores on whatever tool they are using should be identical on the majority of elements/components that are assessed).
- ❖ Administrators responsible for evaluation must identify the primary evaluator for each teacher by September 30 of each school year. The primary evaluator is responsible for providing the teacher with the summative feedback from all observations and developing goals for improvement in consultation with the teacher being evaluated.
- ❖ In addition to receiving the vendor-provided training on the observation tool, the administrator responsible for a teacher’s evaluation must receive adequate training in coaching and providing feedback, including having opportunities for supervised practice. These skills are crucial to the development of an improvement-focused system.

The Mechanics of Observations

- ❖ LEAs have many different types of non-teacher, non-administrator, and ancillary staff. LEAs are responsible for defining and evaluating employees that fall within this category of school personnel.
- ❖ Multiple observations must be conducted for every teacher across the school year. Evidence from the pilot study and from other research indicates that reliable estimates of a teacher's performance depend on at least three observations. Each observation needs to be long enough to gather the information necessary to use the observation tool reliably and with fidelity, but does not need to be a full class period. At least one observation should be unscheduled.
- ❖ Qualified peer observers may observe teachers and provide observational data. A qualified peer observer must be selected by the administrator responsible for a teacher's evaluation. A qualified peer must be a teacher whose performance is rated professional for at least three years in a row and who has additional skills in supporting the development and improvement of practice. During the first three years of implementation of the system, during which time there will be no teachers who officially meet this three-year requirement, LEAs should use their discretion in identifying qualified peers. Qualified peer observers must also be formally trained by the vendor in the use of the observation tool.
- ❖ When possible, at least one of a teacher's multiple observations should be completed by someone who has expertise in the subject matter/grade level or the specialized responsibility of the teacher.
- ❖ Qualified peers may not be the only source for observation data. The administrator responsible for a teacher's evaluation must conduct at least one observation for a teacher's evaluation record and must use results of others' observations (if applicable) in making an overall appraisal of the teacher's practice.
- ❖ Observers (including qualified peers and administrators) should provide teachers with useful developmental feedback on their practice. Useful feedback is specific and framed in ways that provide a practical focus for development.
- ❖ LEAs may use other data that provide evidence about a teacher's practice (e.g., student surveys, parent surveys, portfolios), but for no more than 20% of the practice section.
- ❖ LEAs must decide how to aggregate a teacher's final observation rating from all observations. The state should provide guidance for schools using the state-selected observation tool, as well as data analyses that comply with that guidance.
- ❖ All teacher observation data collected with the state-selected tool shall be reported to the state for research purposes to support continuous improvement in the educator evaluation system.

INFRASTRUCTURE AND IMPLEMENTATION

Observations of teaching might seem simple to carry out. But effective observation is more than just watching teaching and providing comments. Observation that is fine-tuned to key aspects of practice and that is systematic and not opinionated or idiosyncratic requires well-designed checklists and guides, as well as high-quality training for observers. The reliability and validity of inferences drawn from these observations—and hence, their usefulness—depend on the rigor of both the tools and the process. Instruments must have been field-tested and their reliability analyzed. Observers must be properly trained. The processes for observing teachers must be specified so that they are transparent and fair.

Observation Guidelines

As noted, administrators and/or other designated observers must conduct multiple observations each year. Research makes it clear that one observation per year is not sufficient to make a reliable estimate of instructional quality. Too few observations may actually provide inaccurate information. Although there is no definitive answer to how many observations of what length are sufficient, researchers have consistently found that multiple observations lead to higher levels of reliability, and recommend that, when the data will be used for high-stakes evaluation, teachers must be observed during multiple lessons. In order to represent the teacher's practice, the MCEE recommends that observations be conducted at intervals of at least 30 days.

Administrators responsible for evaluations must decide who should evaluate teachers. Observers must be principals, assistant principals, "master teachers," curriculum directors, superintendents, or assistant superintendents. A qualified peer evaluator cannot be the only source for observation data. The administrator responsible for a teacher's evaluation must conduct at least one observation to contribute to a teacher's evaluation record and must be responsible for the final summative evaluation. Every individual who conducts an observation must be trained in using the protocol.

For observations to contribute to the continuous improvement of instruction, teachers must receive productive feedback oriented to the development of their practice. This feedback should be provided through conferences between the observer and the teacher and should reinforce specific strengths and highlight practices or skills that should be developed or improved. Providing targeted useful feedback also requires training; observers should receive professional training in analyzing instruction, providing developmental feedback, and being knowledgeable about relevant resources for teachers to use.

Other Methods of Evaluating Teachers' Practice

Observations are only one component of educator evaluation. Others include documents that support the observations, as well as other materials contributed by teachers, principals, students, or parents. Among the components used in other states are the following:

- Teacher self-assessments;
- Professional development activities;
- Educator growth plans (developed by teachers or administrators);
- Structured review of student work;
- Teacher artifacts using portfolio or evidence binder processes;
- Feedback from students, parents, and/or other teachers using structured survey tools; and
- Teacher self-reflection and progress on professional growth goals.

Although many of these components could be effective in highlighting strengths and areas for improvement in teacher practice, their use is not required. LEAs may use other local data that provide evidence about a teacher's practice, but for no more than 20% of the practice section. The state might play a role in disseminating information about the range of components that can be used and research on their promise and technical merit. The MCEE supports the use of multiple sources of evidence, but is also concerned about the burden this educator evaluation system will place on schools and LEAs, teachers and administrators. LEAs are urged to add components to the system only if the benefit of the additional information is appropriate given the additional cost and effort entailed by its inclusion.

Training and Professional Development

The documentation of teaching is only as good as the observer. One of the pilot study's most significant findings is the importance of training for observers on the evaluation tools. Observers need to be trained to observe carefully, attend rigorously to the key elements of instruction, to be thorough and accurate in their note taking and assessments, and responsible in the conclusions they draw from their observations. Evaluator training increases the reliability and validity of observation tools and enhances administrators'

ability to deliver fair, accurate, and usable feedback to teachers. Therefore, the council recommends that all designated observers undergo the tool vendor's recommended training protocol before conducting any observations. This training regime should be consistent with what is known from research about training requirements for evaluators. This research suggests that those conducting observations should be provided with training in the conceptual framework underlying the observation protocol, training in the meaning of scoring rubrics used in observation, training in the steps that should be taken to conduct an evaluation, practice in scoring the protocol under the supervision of an expert, and training in how to communicate evaluation results to teachers. The state shall make this training available annually for administrators new to an LEA, administrators new to the observation tool, or administrators that need additional support using the tool based on their own administrator evaluation. In addition, the state must make this training available annually for all qualified peer observers. Because there can be drift away from rigorous use of instruments, observers must be retrained every three years. This continued training helps to ensure that observers' use of the tool remains aligned with the tool vendor's standards. LEA personnel should also have access to online resources to support learning and professional development around these evaluation tools and protocols.

Weighting

Weighting is the process of assigning different values to portions of an evaluation to reflect their relative importance in determining overall performance. The MCEE recommends that 50% of a teacher's evaluation be based on the practice component beginning in the 2015–16 school year.

The MCEE recommends that at least 80% of the practice component come from classroom observational evaluation data using one of the four piloted tools. LEAs may use other local data that provide evidence about a teacher's practice (e.g., student surveys, parent surveys, portfolios) for no more than 20% of the practice section.

LEAs may choose to designate the entire practice component of a teacher's evaluation to observation data; or, an LEA may choose to incorporate other methods listed above to complement a teacher's observation data in order to arrive at a total practice rating.

Practice Ratings

In order for teachers to receive a complete evaluation that accounts for all observations over the course of the year, each LEA must establish a method for arriving at a final rating from the multiple observations (as well as any additional information chosen by an LEA) that is reasonably consistent across teachers and time. The MCEE recommends that LEAs be given the discretion to decide how these ratings are processed; however, the local policy must be consistent and publicly available for teachers, administrators, and the public to read.

Teacher Evaluation: Student Growth

One cornerstone of the new educator evaluation system is the use of student growth and assessment data as 50% of an educator's evaluation starting in the 2015–16 school year. "Student growth" refers to the change in students' knowledge and skills across time. The focus in educator evaluation is to try to estimate the changes that can be attributed to the instruction that students receive. One way to do this is to use changes in students' scores on two administrations of a particular assessment, but this does not by itself provide a fair estimate of how much of the change is due to the teaching. Another way to do this, when standardized tests are used to measure student learning, is to use statistical techniques called "value-added" models that are designed to try to isolate the contribution of instruction by controlling for other factors that might impact students' growth. Still another approach is to provide evidence of the students' progress toward a set of articulated learning goals.

During 2013–14 and 2014–15, during which time the new educator evaluation system is not yet fully implemented, the MCEE recommends that LEAs use student growth as a significant component of

teacher evaluations, but not for more than 50% of an individual teacher's evaluation. To prepare for full implementation in 2015–16, it may be advisable for LEAs to pilot the use of student growth and assessment data as 50% of an educator's evaluation during 2013–14 and 2014–15. PA 102 stipulates that the growth and assessment data used for these evaluations will be reliable and valid, measure both core academic subject areas and others, can be used for students across all achievement levels, complies with current laws concerning students with disabilities, and involves at least a pre- and a post-test.

"Student growth and assessment tools" and "value-added modeling" are not interchangeable concepts. Growth and assessment tools are tests that measure the achievement and growth of individual students, and when used on more than one occasion, can assess changes in students' achievement. Value-added models (VAMs) are statistical techniques that use data produced by growth and assessment tools to estimate the effects of teachers (and schools) on their students' achievement. In this report, we explain each in turn.

Although there is much interest in using both student growth and assessment tools and VAMs, considerable scientific concern exists about their instability and measurement error. To ensure care in estimating educators' contribution to student growth, multiple measures should be used to assess student growth. Thus, we recommend the use of a suite of student growth and assessment tools in ways that meet the spirit of the law while also acknowledging the limitations of current assessments and models for establishing an educator's "value-added." We also recommend that the state be vigilant in examining the use of these tools and in tracking their strengths and limitations. We begin with student growth and assessment data before moving to value-added modeling.

STUDENT GROWTH AND ASSESSMENT DATA

The use of student growth—a student's progress between two points in time—in educator evaluations constitutes a high-stakes use of assessment data. Although the concept might be intuitively appealing to many, it is important to understand several core technical issues.

First, there are multiple ways to measure growth. While giving students the same test at time A and time B might seem like the most sensible approach, it involves testing students on material they have never studied, which is educationally problematic. Another approach involves giving students a pre-test based on less complex and less difficult prerequisite knowledge and a post-test based on more complex knowledge and skills. Using well-tested psychometric procedures, experts can create approximately comparable scales for these two tests. Not all experts agree that this is the right approach, however. An alternative is to use statistical models to identify which students progressed more or less than others over the period from the pre-test to the post-test by examining achievement on the post-test controlling for achievement on the pre-test.

Second, assessments useful for measuring student growth can be provided in multiple ways:

- Full-service assessments, which are developed, administered, scored, and reported on centrally, using rigorous professional measurement practices.
- Model assessments, which are developed centrally, using rigorous professional measurement practices, but administered, scored, and reported on locally.
- Locally developed assessments, which are developed, administered, scored, and reported on locally.

Another way to evaluate student growth is through the adoption of student learning objectives (SLOs), or specific measurable academic goals that teachers and evaluators set for groups of students. Student learning objectives may be shared by a team of similar teachers across a state, LEA, or school. Several states and large LEAs currently use a specific approach to identifying student learning objectives to measure student learning in non-tested grades and subjects. In order for SLOs to yield useful information, they must be appropriately rigorous, cover a reasonably representative portion of the curriculum, and be accompanied by tools to measure students' progress toward goals. This requires training so that teachers

and administrators can create, implement, and assess student growth on SLOs in ways that ensure fair standards and high quality.

Given the current state of available assessments, requiring high-stakes use of assessment data obligates the state to provide the following high-quality tools at state expense:

- Assessments aligned to state-adopted content standards in all core content areas (i.e., English language arts, mathematics, science, and social studies);
- Assessments aligned to state-adopted content standards in high-volume non-core content areas where state-adopted content standards exist (e.g., arts, health and physical education, career and technical education, and many high school electives);
- Guidelines for minimally required evidence of appropriate quality for assessments provided by third parties;
- Guidelines for minimally required evidence of appropriate quality for locally developed assessments; and
- Guidelines for training on adequate rigor in the development and measurement of the achievement of SLOs.

Any state-provided assessments should produce results that are useful in educator evaluations in a timely fashion. Therefore, the state must:

- Provide growth data back to LEAs for individual students on full-service yearly assessments; and
- Provide growth data back to LEAs aggregated for individual educators.

Educator evaluation using student growth data is a field still under considerable development. Given this, it is crucial that Michigan educators are given some flexibility in the adoption of tools for measuring growth while also maintaining rigor. Therefore, the MCEE recommends that:

- As it relates to educator evaluation, the state should not mandate the adoption of additional assessments beyond those required by federal and current state statute;
- LEAs should be allowed to adopt additional assessments provided by the state or by third parties if they meet basic quality requirements;
- LEAs should be allowed to adopt locally developed assessments if they meet basic quality requirements; and
- In the event that the state revises or adopts new content standards in any content area, new assessments should not be used in evaluations for at least two years after the adoption of the new content standards to assure opportunity for new item development and educator professional development.

As we have noted, the MCEE is committed to creating a system of educator evaluation that is maximally educative for all. It is our obligation to focus educator and student attention on **learning**, and any and all assessments that are used should also provide information to teachers and administrators about areas of improvement. Thus, the MCEE recommends that any and all high-stakes assessments and growth tools that are used have the capacity to provide educators with useful information about student knowledge and performance that is more specific than a scaled score, including item-level information when technically or logistically possible, so that teachers and administrators can use the data to inform their ongoing work. This can include strand-level data (e.g., “number sense and operations” as a subset of mathematics), data segmented for particular learning groups (e.g., student growth data on students below grade-level, special populations, etc.), and test score histories for the students for whom the teacher has instructional responsibilities.

In addition, if there are changes in the state content standards, the state will need to develop new assessment items, pilot test those new items while also using the old items, and develop and offer professional development for educators in the changes in content standards and assessment. It typically

takes about two years before there are new assessment items and broad understanding among educators about the ensuing changes in standards and assessments of student learning and growth.

Growth and Assessment Tool Recommendations

Given this array of issues, the selection of appropriate growth and assessment tools, aligned with state standards for students' learning, is of critical importance. LEAs should use the tools provided by the state for the subjects and levels that are part of the state assessment system. In subjects and at levels for which no assessments are provided, LEAs must select appropriate tools or approaches to measure student growth.

The tables beginning on page 18 show the options from which LEAs will be able to choose, beginning in the 2014–15 school year. Represented are tools to measure student growth in all subject areas in grades K–2, 3–8, and 9–12.

The title of each table describes the grade band. The first column lists the subject area. The other columns show the different sources of assessments from which LEAs may choose to provide measures of student growth. These are separated into five sources.

The first column representing sources of assessments is for standardized assessments paid for by the state, required for all public school students. These assessments are typically administered only once per year. In this column, “A” or “a” indicates that the assessments provided by the state are fully aligned to the content standards adopted by the state. These represent assessments that are known to measure the content that students are supposed to learn and that educators are supposed to teach. An uppercase letter in this column represents assessments given throughout the full span of grades, with a lowercase letter representing assessments given only in select grades in the span. For instance, state-provided required assessments in science and social studies only have to be given once covering content required in grade spans 3–5 and 6–8. Additionally, students are only required to take the full high school assessment covering mathematics, reading, science, social studies, and writing once during high school, typically in grade 11. Even those assessments that are not provided throughout an entire grade span can be used as a post-test for a different assessment covering the same content area.

These required state-provided assessments may be either developed specifically to match state content standards, or developed independently by vendors and shown to match state content standards.² Verification of alignment to the state content standards will be made by the state using an independent contractor. It is not sufficient to use a vendor's claim as evidence that an assessment tool matches the state content standards. For these assessments, the state should be required to provide value-added measures back to LEAs, for each educator responsible for the education of individual students.

The next column representing sources of assessments is for standardized assessments paid for by the state, but adopted optionally by LEAs. In this column, an “A” indicates a full-service assessment known to align fully to the content standards adopted by the state, with the cost of development, scoring, and reporting being paid by the state. An M indicates a model assessment known to align fully to the content standards adopted by the state. These assessments are developed centrally by the state, but are administered, scored, and reported on locally. These assessments could also be administered annually or more often. These assessments are important in order to make possible:

- Comparability of growth measures across LEAs (to the extent that LEAs adopt them).
- Use of high-quality assessments developed using professional practices.
- Use of assessments known to measure the content students are supposed to learn and educators are supposed to teach.
- Relief for LEAs of the expense of developing their own assessments and/or adopting independent commercial assessments.

² Federal law requires states to use assessments that are aligned with the state content standards.

Again, these optional state-provided assessments may be either developed specifically to match state content standards, or developed independently by vendors and verified by the state to match state content standards. For optional assessments that are developed, administered, scored, and reported on centrally by the state, the state should be required to provide value-added measures back to LEAs that adopt the assessments, for each educator responsible for the education of individual students. LEAs should be allowed to determine whether to include those value-added measures in their local evaluation systems. The MCEE does not recommend that such assessments be provided by the state in the specific content areas and grade levels identified in the table. These are provided as illustrative content areas based on the Michigan Department of Education's (MDE's) current plans for development and/or adoption of assessments in service to educator evaluation. Assessments in this column (providing high-quality measures of growth in all core content areas) will require new ongoing appropriations in order to support educator evaluation adequately.

The next column indicates assessments provided independently by vendors. An "X" in this column indicates that it is known that such assessments exist for the specified content area, but their alignment to state-adopted standards is unknown. Finally, a question mark in this column indicates that we are unaware of vendors providing standardized assessments in these content areas, and that such assessments may or may not be developed in the future.

It is important that LEAs be able to adopt assessments outside the suite of assessments paid for by the state. However, because of the unknown alignment of vendor-provided assessments to the state standards, any LEA that adopts such an assessment should put into place policies that clearly lay out expectations concerning the content to be taught to students. That is, if an LEA expects educators to teach content above and beyond state content standards, those expectations need to be clearly articulated and explicit. Also, because of unknown alignment of vendor-provided assessments to official state standards, these assessments should not form the sole basis for measuring student growth for educator evaluations. Any LEA's adoption of vendor-provided assessments should require a demonstration of adequate reliability and validity of the assessments before using scores from those assessments in educator evaluations.

The next column indicates assessments developed and adopted locally at the ISD or LEA level. A checkmark in this column indicates that LEAs should be allowed to adopt such assessments in the specified content area. Locally developed assessments adopted for measuring student growth for use in educator evaluations should be developed specifically to align fully to the official state content standards where official state content standards exist. In addition, any LEA's adoption of locally developed assessments should require a demonstration of adequate reliability and validity of the assessments for inferences about student growth and teacher evaluation before using scores from those assessments in educator evaluations.

The last column indicates grades and content areas in which the use of student learning objectives (SLOs) that measure a sufficiently representative sample of the subject area should be allowed in measuring student growth for inclusion in educator evaluations. When SLOs are adopted as a measure of student growth by an LEA, adequate training in the development and implementation of SLOs should be provided.

ASSESSMENTS TO BE AVAILABLE IN 2014-2015

Key	
A	Assessments are fully aligned to the content standards adopted by the state, are developed, administered, scored, and reported on centrally by the state, and are given throughout the full span of grades.
a	Assessments are fully aligned to the content standards adopted by the state, are developed, administered, scored, and reported on centrally by the state, and are given only in select grades.
M	Assessments are fully aligned to the content standards adopted by the state, are developed centrally by the state, but are administered, scored, and reported on locally.
X	Assessments are developed by vendors but their alignment to the content standards adopted by the state is unknown.
?	We are unaware of vendors providing standardized assessments in these content areas, and such assessments may or may not be developed in the future.
✓	Districts should be allowed to adopt assessments or develop SLOs in the specified content area for measuring student growth. These should be developed specifically to align fully to the official state content standards where official state content standards exist.

Grades K–2: Subject Areas and Assessments

Subject Area	Assessment Sources				
	State Standardized		Vendor Standardized	ISD/LEA Developed	Local SLOs
	Required	Optional			
Mathematics		A	X	✓	✓
Reading		A	X	✓	✓
Science			X	✓	✓
Social Studies			X	✓	✓
Writing			X	✓	✓
Art		M	?	✓	✓
Music		M	?	✓	✓
Physical Education			?	✓	✓
Foreign Language			X	✓	✓
Other			?	✓	✓

Grades 3–8: Subject Areas and Assessments

Subject Area	Assessment Sources				
	State Standardized		Vendor Standardized	ISD/LEA Developed	Local SLOs
	Required	Optional			
Mathematics	A	A	X	✓	✓
Reading	A	A	X	✓	✓
Science	a	A	X	✓	✓
Social Studies	a	A	?	✓	✓
Writing	A	A	X	✓	✓
Art		M	?	✓	✓
Music		M	?	✓	✓
Physical Education			?	✓	✓
Foreign Language			?	✓	✓
Other			?	✓	✓

Grades 9–12: Subject Areas and Assessments

Subject Area	Assessment Sources				
	State Standardized		Vendor	ISD/LEA	Local
	Required	Optional	Standardized	Developed	SLOs
Mathematics	a		X	✓	✓
Algebra I		A	X	✓	✓
Algebra II		A	X	✓	✓
Geometry		A	X	✓	✓
Other math			X	✓	✓
Integrated math			?	✓	✓
English	a		X	✓	✓
English 9		A	X	✓	✓
English 10		A	X	✓	✓
English 11		A	X	✓	✓
English 12		A	X	✓	✓
Other English			?	✓	✓
Integrated English			?	✓	✓
Science	a		X	✓	✓
Biology		A	X	✓	✓
Physics		A	X	✓	✓
Chemistry		A	X	✓	✓
Other science			?	✓	✓
Integrated science			?	✓	✓
Social Studies	a		X	✓	✓
Civics		A	X	✓	✓
US History & Geography		A	X	✓	✓
World History & Geography		A	X	✓	✓
Economics		A	X	✓	✓
Other social studies			?	✓	✓
Integrated social studies			?	✓	✓
Arts			X	✓	✓
Visual arts		M	X	✓	✓
Performing arts		M	X	✓	✓
Applied arts		M	X	✓	✓
Other arts			?	✓	✓
Physical Education			?	✓	✓
Foreign Language			X	✓	✓
Other			?	✓	✓

To maintain quality and ensure a continuous improvement program of state assessments, the state should be given authority to:

- Monitor compliance with legislative requirements for measuring student growth; and
- Conduct an evaluation of the quality of locally adopted tools for measuring student growth to inform potential improvements in requirements over time.

The ongoing evaluation of the quality of all growth and assessment tools used in the state is crucial. Requiring the insertion of this level of assessment into schools, and using these assessments for high-

stakes decisions about educators' jobs, makes it imperative that the state be vigilant both about the quality of the tools being used and new developments in the field of growth and assessment tools more generally.

VALUE-ADDED MODELS

As mentioned above, value-added models (VAMs) are separate from growth and assessment tools. VAMs are statistical models that use data from growth and assessment tools to produce estimates of the "value added" by individual educators to student learning. VAMs attempt to isolate the effects of individual educators on the achievement or growth demonstrated by each educator's students by "leveling the playing field" for educators. This is done by controlling for factors over which educators have little to no influence. Some such factors include individual students' incoming achievement, race/ethnicity, socioeconomic status, gender, special education status, and English language learner status. Other factors over which individual educators have little to no control may be aggregate factors, such as the average incoming achievement of a teacher's classroom, the level of variability in incoming achievement in a principal's school, the average socioeconomic status of a teacher's classroom, or the proportion of English language learners in a principal's school. In addition, the statistical techniques used in VAMs are designed to minimize the risk of misclassifying teachers based on their students' test score gains, which is particularly important for those with small classes.

VAMs may be based on measures of student growth, but do not need to be. Measures of value added for an individual educator are based on the typical deviation of his or her students' achievement or growth from the achievement or growth those students were expected to demonstrate given previous achievement and/or other factors over which the educator has little to no influence.

There are many different approaches to measuring the "added value" of an individual teacher's impact on students' growth or achievement, but there is scholarly disagreement over the appropriateness of these various approaches. Some researchers are skeptical about VAMs in general because they question the validity of making causal claims about the impact of individual educators on student outcomes, especially given that students are not randomly assigned to teachers and schools. Others argue that VAMs should not be used for high-stakes decisions because of the significant measurement error inherent in them. Others assert that the use of VAMs is appropriate, but that in order to be fair to educators, it is necessary to control for all factors (for which there are data available) over which educators have little to no influence. Finally, others assert that the use of VAMs is appropriate, but that in order to be fair to students, VAMs should not control for demographics because doing so implicitly sets different expectations for student growth and/or achievement based on demographics, resulting in expectations that achievement gaps will persist and possibly become larger.

Given the scientific disagreement about VAMs, some policymakers may question their use in high-stakes educator evaluation. The MCEE recognizes the well-documented limitations of these models, and cautions against their overuse in educator evaluation. However, when comparing the use of VAM data to the alternative of district-developed data models of teaching effects, the MCEE believes that VAMs provide more reliable evidence. The MCEE's review of student data practices in Michigan and other states revealed that VAMs include many more statistical controls than what the vast majority of LEAs are able to build on their own, use more reliable assessment data, and involve more deliberate efforts to control for data quality issues that may bias the data. VAMs, while far from perfect, represent a better and fairer alternative to other means of ascribing test scores to educators.

VAMs should only be based on assessment data that provide information that support valid and reliable inferences about students' growth. In subject areas for which there are no VAM data available, educators should be evaluated based on alternate measures of student growth and achievement.

Multiple VAM measures will be available for some educators because: (1) VAMs are typically run one content area at a time while some educators teach more than one content area; and (2) VAMs are typically calculated separately for each section for which an educator is responsible and some educators

teach more than one section. The method for combining multiple VAM measures for a single educator should be left to local discretion.

The state should provide VAM scores for individual educators on all state-mandated assessments and all optional assessments offered by the state.

KEY VALUE-ADDED RECOMMENDATIONS

- ❖ For teachers in core content areas in grades for which there are growth data available from state-mandated assessments (currently reading and mathematics in grades 4–8, but likely to change over time), at least half of the teacher’s student growth component should be based on state-provided value-added estimates.
- ❖ For teachers of other subject areas, LEAs should have the discretion to adopt state-provided VAM or growth data.
- ❖ When possible, state-provided VAM or growth data should be based on three-year averages to reflect persistent teacher effects.
- ❖ State-provided VAM or growth data in core content areas may be used in a teacher’s evaluation for that teacher’s students, even if the teacher does not teach in one of the core content areas. This means that teachers may be evaluated, in part, for the learning of their own students, even in subject areas that they do not directly teach. This may be done as long as the teacher knows that he or she is expected to be contributing to students’ growth in core content areas and there is a reasonable connection of the core content to the teacher’s actual teaching assignment. For example, a high school history teacher’s evaluation may include her students’ growth in English language arts because she is responsible for helping her students develop as readers and writers within her content area.
- ❖ In order to promote collective work on instruction within school buildings, school-level VAMs may be used for individual teachers’ evaluations if there is a reasonable connection of the core content to the teacher’s actual teaching assignment. This means that teachers may be evaluated, in part, for the learning of students whom they do not directly teach. However, school-level VAMs may not comprise more than 10% of the individual teacher’s student growth component.
- ❖ In content area assignments for which there is no state-provided VAM or growth data available, teachers should be evaluated based on alternate measures of student growth that meet the guidelines for rigorous and appropriate assessment of student growth.
- ❖ There are complications in calculating VAM that will need to be addressed in designing models to estimate the effects of individual educators on student learning. These include:
 - Determining a method to appropriately apportion responsibility for a specific student’s data to multiple educators in situations where more than one educator is responsible for the education of that student in a single content area, such as in team teaching.
 - Determining whether and how to control for multiple student and class, school, or LEA characteristics.
 - The legislative requirement that an educator be able to appeal to have any specific student removed from his/her VAM score.

These complications make it necessary to make adjustments to state data systems to capture fine-grained attendance data for individual students, proportional responsibility by student for educators, and a roster review and appeal mechanism with approval by supervising administrators. Because the data linking students to individual educators will now be used in a high-stakes manner, it is imperative that individual educators be able to review those data links for accuracy, and to appeal inaccurate links, approvable by the educators’ supervising administrator. Sufficient appropriations should be made to support the enhancement of state data systems to provide for these functions.

- ❖ State-produced VAMs should provide mechanisms that allow for exclusion of students whose total attendance is less than 90% of the school year. Total attendance should be calculated based on the total number of school days a student misses and not on his or her attendance in the specific class from which the teacher's VAM is derived. Students' assessment data may also be excluded based on an educator's appeal and supervisor's approval to have any specific student removed from his or her VAM score.
 - ❖ In addition, because VAM is a separate concept (and activity) from measurement of student growth and achievement, a separate vendor will need to be procured to carry out the VAM work. Given the limitations of value-added models, it is especially important that the vendor make every effort to mitigate those limitations. This contractor should be required to:
 - Contract with the state to carry out all VAM activities.
 - Receive data from the state in a specified format.
 - Develop, in collaboration with relevant stakeholders, the requirements for building and estimating VAMs, while considering the conflicting opinions about the appropriateness of including demographics in the VAMs, and reporting VAM results back to educators. This should include:
 - Requiring VAM estimates to be reported with standard errors;
 - Requiring the decision rules for using the estimates to reflect those standard errors; and
 - Publishing regular analyses to demonstrate how many teachers' VAM estimates vary year by year.
- Or:
- Requiring estimates to be based on a minimum number of student data points; and
 - Requiring estimates to be based on a minimum number of separate classrooms (to make it possible to separate teacher effects from classroom effects).
 - Develop, in collaboration with the state, the requirements for feeding data back to state systems, developing the VAM reporting portal in a manner compatible with state requirements, and providing secure access to individual educators' VAM reports through state systems.
 - Provide a clear explanation of the VAM method used that is understandable to the general public.

Combining and Implementing the Practice and Student Growth Components

The MCEE recognizes both the complexity and importance of considering how data produced from observational and student growth tools are combined into a final evaluation rating. The MCEE has reviewed the approach for combining evaluation data used in states such as New York, Ohio, Tennessee, North Carolina, and Colorado.

In some systems, states provide weights to various components of a teacher evaluation system and use a formula to generate a final combined score. The primary advantages to this approach are its flexibility in the number of inputs included (many measures can easily be integrated: teacher observation data, student growth data, parent satisfaction data, etc.) and its flexibility in weightings (the system can easily apply different weights to each component, depending on priorities). This approach has two primary disadvantages. First, it has the potential to communicate a false degree of precision given the error commonly found in teacher observation data, student growth data, and value-added data. Second, this approach also assumes that each component is placed on comparable scales before combining them. Placing diverse data elements like value-added estimates and teacher observation data on a common scale has complications that limit the utility of this approach.

In other systems, teacher observation data and student growth data are both independently mapped to standards of performance. The MCEE favors this approach for three primary reasons. First, this approach has the potential to help manage imprecision inherent in student growth scores and observation data by classifying both into general ranges. Second, the means of combining scores is clear and transparent. Mapping data from the categories of performance to teaching effectiveness ratings makes it clear to those being evaluated, doing evaluations, or using evaluation data how the effectiveness ratings are created. Third, this system allows adequate control when input data such as student growth scores or observational data are unusually high or low. For example, if evidence from a teacher's observations is strong, but data show weak student growth, what should be that teacher's evaluation rating? These systems for combining scores allow for specific decision logic for these situations, and those decisions will not bias ratings created from more typical outcomes.

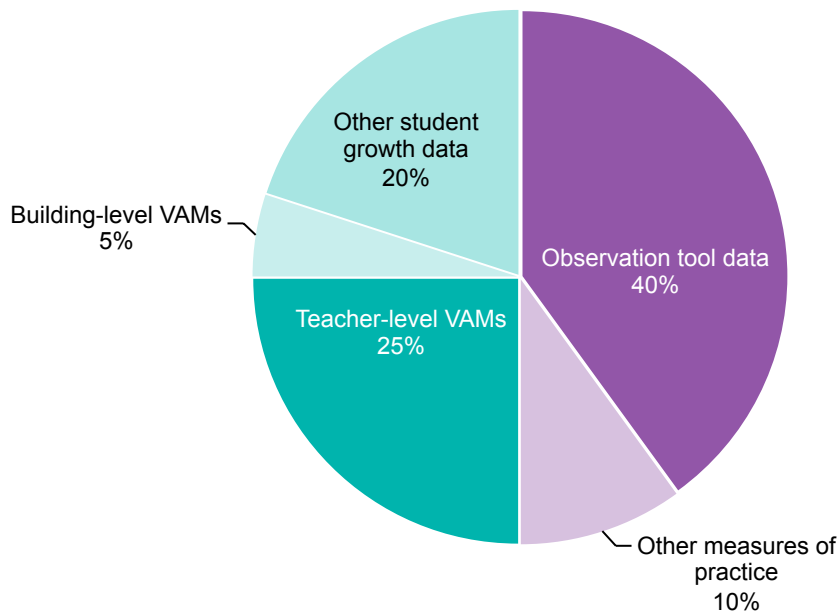
KEY RECOMMENDATIONS FOR COMBINING SCORES

- ❖ During the first two years of the teacher evaluation system (i.e., 2015–16 and 2016–17), LEAs will produce teachers' ratings based on qualitative combinations of the categories. Results of teachers' observations will be ordered in terms of skill and proficiency, and VAM or growth scores will also be ordered according to the impact on students' learning. In this interim approach, the categories will represent an ordering in terms of overall quality, but the differences do not have the properties of numerical scales. One possible example of a set of evaluation rankings produced by considering both the quality of the teacher's practice and student growth is shown below:

		Practice		
		Professional	Provisional	Ineffective
Student growth	Meets expectations	Professional	Professional	Provisional
	Does not fully meet expectations	Professional	Provisional	Ineffective
	Below expectations	Provisional	Ineffective	Ineffective

- ❖ During the first two years of the new system, a standard-setting process should be conducted based on data from the teacher evaluation system so that, in the longer term, it will become possible to produce profiles of teachers' performance based on specific aspects of instructional quality and student progress. In this way, teachers' actual instructional proficiency will become the basis for their evaluation and feedback, rather than simplistic numerical scores.

One example representation of the proportional relationship of the types of data that can contribute to a teacher's overall rating beginning with full implementation in 2015–16 is on the following page. This example applies to teachers in core content areas in grades for which there are growth data available from state-mandated assessments and assumes the minimum use of teacher-level VAMs (50% of the student growth section; 25% of the overall evaluation), the maximum use of building-level VAMs (10% of the student growth section; 5% of the overall evaluation), and the maximum use of other measures of practice (20% of the practice section; 10% of the overall evaluation).



Administrator Evaluation Recommendations

OVERVIEW

Administrators play a central role in high quality instruction. They support teachers, provide feedback, and enable and enhance professional learning communities. They manage resources, work with parents, and play a critical role in communicating between the school and the community. As with teacher evaluations, the MCEE also recognizes the importance of having fair, transparent, and feasible administrator evaluation tools.

BACKGROUND

The MCEE began its work on potential tools and processes for the administrator evaluation by forming a subgroup of council members who reviewed:

- National Board Standards for Accomplished Principals;
- Interstate School Leaders Licensure Consortium (ISLLC) Standards for School Leaders;
- Michigan School Code provisions and the language of the new legislation;
- National Association of Secondary School Principals (NAESP) and National Association of Elementary School Principals (NAESP) Principal Evaluation;
- Research and research syntheses on administrator evaluation;
- Sample tools used by other states; and
- Sample systems developed by Vanderbilt, Marzano Research Laboratories, Kim Marshall, Mid-continent Research for Education and Learning (McREL), and Dr. Douglas Reeves (The Leadership and Learning Center).

The MCEE also asked the Advisory Council to review materials and make independent recommendations based on their review. Using criteria that were developed by the Advisory Council with advice from the authors of the ADvance materials (Reeves and McNeil), the Advisory Council recommended Marzano, School ADvance, and New Leaders for New Schools.

The MCEE engaged the Survey Research Operations Group of the Survey Research Center at the Institute of Social Research to conduct ten focus groups across the State of Michigan. The purpose of the focus groups was to gather qualitative data regarding five administrator evaluation tools selected by the MCEE. Participants in the focus groups were superintendents or individuals in other roles with responsibility for evaluating school principals. The five tools reviewed were:

- The Marzano School Leadership Evaluation Model;
- MASA's School ADvance Administrator Evaluation Instrument;
- Reeves Leadership Performance Rubric;
- Illinois State Model for Principal Evaluation; and
- Indiana (RISE) Principal Effectiveness Rubric.

See Appendix F for links to all of these tools and systems and Appendix G for the full report from the Survey Research Operations Group at ISR.

KEY RECOMMENDATIONS

- ❖ LEAs must choose from one of two administrator evaluation tools:
 - MASA's School ADvance Administrator Evaluation Instrument; or
 - Reeves Leadership Performance Rubric
- ❖ One of these tools will be selected to be the state tool. The state-selected tool will be chosen after a competitive RFP process. The state will provide sufficient base funding per administrator to support LEAs' use of the state-selected tool with full fidelity. The state will also provide the technical support and training for the state-selected tool.
- ❖ LEAs may choose to use the other tool instead. An LEA that chooses to use the non-state-selected tool must pay for any expenses above the base funding supplied by the state.
- ❖ Like teachers, administrators must be evaluated on both their practice and student growth measures.
- ❖ According to the law, administrators (e.g., principals, assistant principals, curriculum coordinators, superintendents, assistant superintendents, career and technical education managers, special education directors) must be evaluated by at least the following evidence for each school (or LEA, for superintendents or central office personnel):
 - Proficiency of skill in evaluating teachers;
 - Progress made in the school improvement plan;
 - Attendance rates; and
 - Student, parent, and teacher feedback.

Additionally, LEAs may choose to incorporate other pieces of evidence that demonstrate an administrator's professional effectiveness and growth. Some examples may include, but are not limited to, professional contributions, peer input, and training/professional development.

- ❖ The school board must evaluate superintendents.
- ❖ The state should provide hosting and data analysis services for the state-selected tool for all Michigan LEAs that adopt that tool.

Key Steps in the Process

In order to provide administrators with the type of formative feedback that teachers receive when they are observed and have opportunities to discuss their observations, the MCEE recommends the following:

- ❖ Administrators and their supervisors should meet each year before October 1. As part of this planning, school principals and their supervisors will establish an agreement about how information on (1) the proficiency of skill in evaluating teachers; (2) progress made in the school improvement plan; (3) attendance rates; and (4) student, parent, and teacher feedback will be presented during the year so that that information can inform the summative decision at the end of the year. Other administrators will discuss how comparable information will inform the process.
- ❖ By February 1 of each year, administrators and their supervisors should have a mid-year evaluation conference in which supervisors will provide administrators with verbal and written feedback (including relevant information on teacher evaluations, student and parent feedback, attendance rates, and school improvement work), including clear information on any areas of concern that should be addressed by June.
- ❖ At the end of the school year, supervisors and administrators will have a summative evaluation meeting. Supervisors will provide administrators with both verbal and written feedback, and provide the administrator's final evaluation rating using both the selected rubric and the information provided on teacher evaluations, school improvement progress, attendance, and student, teacher, and parent feedback.
- ❖ Administrators' performance will be rated using the same three categories as teachers: professional, provisional, and ineffective:
 - **Professional:** An administrator rated **professional** has exhibited the knowledge and capabilities expected of a skillful leader. We expect that administrators who are extraordinary—as well as administrators who are competent—would fall into this category. We also assume that even highly skilled practice can be improved. Because the purpose of educator evaluation is improvement, and because all educational practice benefits from ongoing refinement, a rating of “professional” should always be accompanied by specific feedback for development. An administrator rated as professional for three straight years may pursue opportunities for advanced roles or leadership. In addition, an administrator rated as professional for three straight years may be evaluated on an every other year basis in subsequent years and receive two-year goals for enhancement.
 - **Provisional:** An administrator rated **provisional** has exhibited some professional knowledge and skill, but has specific substantial identified weaknesses that should be addressed through feedback and targeted professional development. We expect that this category might include two groups of administrators: beginning administrators who have yet to develop the level of skill necessary for professional practice and experienced administrators who struggle with mastering professional knowledge and skills. This rating is intended to be a serious signal that the administrator's practice requires significant improvement in specific areas and the rating should be accompanied by clear feedback about areas that need focused attention. An administrator rated as provisional for three straight years should be counseled out of his or her current role.
 - **Ineffective:** An administrator rated **ineffective** has exhibited performance that has specific critical identified weaknesses. The administrator should be placed on urgent notice that significant improvement must be achieved in the specific areas of underperformance within two years. An administrator who receives an ineffective rating for two years in a row should be terminated from further employment as an administrator in his or her current LEA.
- ❖ LEAs will develop a system to manage challenges to an administrator's evaluation.

Calculating Ratings

Administrators will receive ratings of professional, provisional, or ineffective. The legislation stipulates that administrator evaluations use two subcomponents: one for practice and one for student growth. According to PA 102, the practice and student growth subcomponents of an administrator's evaluation will each make up 50% of an administrator's evaluation starting in the 2015–16 school year. During 2013–14 and 2014–15, during which time the new educator evaluation system is not yet fully implemented, the MCEE recommends that LEAs may use student growth as a significant component of administrator evaluations, but not for more than 50% of an individual administrator's evaluation. To prepare for full implementation in 2015–16, it may be advisable for LEAs to pilot the use of student growth and assessment data as 50% of an administrator's evaluation during 2013–14 and 2014–15.

The practice rating will be determined at the summative, end of year meeting, using the LEA's chosen rubric. Assessments within the rubric will be informed by the supervisor's use of the range of evidence.

Student growth data for administrators can be drawn from several sources of evidence. At least half the student growth portion of each administrator's evaluation must come from a building-level VAM scores where available (currently in reading and mathematics in grades 4–8, but likely to change over time). Both reading and math VAM scores must be weighted equally to make up the final building VAM score.

The other portion of the student growth component will be determined at an LEA level. Other examples of student growth data that could be included in this portion of the student growth component are, but should not be limited to:

- Student learning objectives;
- Graduation rates;
- Local common assessment performance;
- State-provided assessments in other content areas;
- Vendor-provided assessments in any content area;
- Locally developed assessments in any content area;
- Pass/fail rates; and
- Percentage of students on track to graduate.

As with measures of growth for teachers, any vendor-provided or locally developed assessments should be adopted for administrator evaluation only after the same safeguards are in place as required for growth measures to be used for teacher evaluation. LEAs must post the policy about what information will be used for an administrator's evaluation and how the sources of evidence will be weighted for all teachers, administrators, and the public to view. As with teacher evaluations, administrator evaluations will only be as good as the training that supervisors receive. Any administrator or school board member who conducts administrator evaluations must be properly trained; retraining will be required every three years, as research accumulates concerning the effectiveness of these evaluation systems and school policies change. It will be the responsibility of the state to determine training protocols for the two selected administrative rubrics.

As with teachers' evaluations, final judgments will be determined by combining student growth scores with practice scores. LEAs will report the summative decision concerning an administrator's assessment to the state.

Administrators will receive their evaluation ratings based on evaluation tool data, available evidence, and other local measures before the end of June. Most likely, state-run VAM estimates will not be released until August, and administrators will receive their full evaluation ratings when those data are available. If decisions about personnel need to be made before the VAM data are ready, then supervisors will use as much evaluation data as is available. We return to these issues in our discussion of the infrastructure necessary to support the new educator evaluation system.

Professional Teaching Certificate

OVERVIEW

Currently, Michigan teachers initially receive provisional certification for the first five years of their career, with an opportunity to renew the provisional certification for three additional five-year terms if necessary. This initial license to teach is labeled “provisional certification.” The next level of certification is called “professional.” Current legislation requires that teachers complete 6 credit hours or 180 hours of professional development to move from the provisional to professional certification. Current legislation is silent on linking teacher evaluation to the granting of the first professional certification.

KEY RECOMMENDATIONS

- ❖ The MCEE believes that it is essential for the state to begin linking professional certification to evidence of effective teaching. Therefore, the MCEE recommends that teachers wishing to move from a provisional to a professional certificate must also receive a professional rating on their annual evaluation for three successive years immediately prior to applying for professional certification. Teachers who do not demonstrate three successive years of professional ratings can apply for the renewal of their provisional certification until they either:
 - Achieve three successive years of professional ratings immediately prior to applying for their professional certification; or they
 - Have three non-successive years of professional ratings but have the recommendation of their current principal.
- ❖ The MCEE also recognizes that building a high quality teacher workforce entails supporting and acknowledging advanced certification for teachers who have steadily demonstrated highly skillful teaching and/or have taken on significant leadership roles as model teachers, mentors for new teachers, or leaders of professional development. Some of these new roles are implicated in the state’s educator evaluation system, as experienced teachers will be a critical component in helping new teachers develop the professional knowledge and skill necessary to achieve a professional rating. Here too the MCEE recommends that only those teachers who consistently receive professional ratings should be eligible to receive advanced certification. However, taking on new roles—mentor, model, guide, or leader—also requires the development of additional knowledge and skill. High teaching evaluations should not be understood to qualify an educator for other kinds of work, including teaching or leading adults.

LEA Waiver Processes and Principles

OVERVIEW

Although the MCEE is committed to creating a quality system of educator evaluation, the council realizes that some LEAs have used time and resources to develop their own evaluation systems. LEAs that are able to demonstrate the effectiveness of their local evaluation tools and policies are welcome to apply for a waiver from the state system.

KEY RECOMMENDATIONS

- ❖ In order to receive a waiver from using the state recommended process, entities requesting a waiver must demonstrate that their processes and systems will have the same level of quality and rigor as those adopted in LEAs following the state requirements. Thus the materials required for the waiver must include the following:
 - A thorough description of a well-articulated evaluation process for teachers, administrators, and other certified staff. For teachers, this will include specifications for the frequency and length of observations that ensure multiple observations and a description of all other performance indicators that educators will submit.
 - Rubrics that allow for detailed descriptions at each level of performance for each indicator (e.g., demonstrating knowledge of content and pedagogy, designing assessments of student learning, managing student behavior, engaging students in learning, communicating with caregivers). An effective rubric will provide meaningful descriptions—not just a number rating—ensuring that teachers and administrators receive detailed, actionable feedback from their evaluators. This includes clear expectations for classroom practice and administrator behavior.
 - A plan and process for giving feedback, including remediation plans.
 - A process for training evaluators on all aspects of the evaluation system, including each piece of the observation tool for teachers, other performance indicators, and the evaluation systems for teachers and other staff as a whole.
 - A process for tracking, managing, and/or importing all data and documentation collected for the evaluations, including observation data for teachers, other information/data, and student growth data.
 - A process for determining summative ratings for all relevant measures (including observation data for teachers and performance indicators for other certified staff), which will then be combined with student growth data as per state requirements.
 - A plan to offer additional direct support to new and struggling teachers or staff (e.g., additional observations, coaching, mentoring).
 - A system for monitoring the fairness, consistency, and objectivity of the system within and across local schools, including specific metrics to be used. For example, LEAs should consider how the distribution of ratings compares with teacher observation ratings and student growth data.
- ❖ If an LEA submits an adapted form of a commercial evaluation system, the LEA must demonstrate how the adaptations do not threaten the validity of the inferences based on the use of the instrument.
- ❖ If an LEA is using an evaluation system that does not have documentation about its validity and reliability, the LEA must submit a plan for how it will gather relevant data on the system's technical soundness. LEAs will receive probationary approval for their waiver, contingent on providing evidence of their system's reliability and validity within three years. (See Appendix H for details about validity and reliability.)
- ❖ LEAs that are denied a waiver can reapply the following year. The committee that will review applications will do so annually with one deadline for requests.

- ❖ An LEA may request a waiver for either the administrator evaluation system, the teacher evaluation system, or both.
- ❖ To put this process in place, the state will need to create a dedicated office for coordination of the waiver process, including:
 - Developing materials (application materials, FAQs, scoring rubrics for plans, model materials).
 - Maintaining a website with materials.
 - Establishing a clear timeline for material submissions that allows reasonable time for review and any necessary response or revisions from LEAs.
 - Communication and capacity building with LEAs and other relevant entities.
 - Record keeping of submitted plans and decisions.
 - Written feedback in each area.
 - Coordinating the review and approval of waiver requests (including clearing probationary cases when reliability and validity evidence is provided).
 - Potentially supplying support or technical assistance to LEAs.
 - Authority to approve and deny waivers, based on rigorous, consistent review of LEA-provided documentation.

Developing, Implementing, and Revising the Evaluation System

BUILDING A LASTING INFRASTRUCTURE

Clearly, creating an educator evaluation system will require developing the relevant policies and processes and creating or adopting relevant instruments to be used. But this system will also require building capacity—both statewide and centrally—to understand, implement, and monitor the system, as well as infrastructure to house data, review materials, produce reports, and conduct program evaluation. In order to carry out these important functions, adequate staffing and resources will be required at both the state and local levels. In Appendix I, we include two sample organizational charts to illustrate how staffing and resources could be structured to meet the extensive demands that this new educator evaluation system will create. Regardless of how it is structured, a framework for monitoring and improving the system should be built in from the start. In order to ensure this, the MCEE recommends the following:

- ❖ The state should be made responsible for monitoring compliance with the law and evaluating the quality of local implementation of the law.
- ❖ The state should be provided with adequate resources to monitor (on a random sampling basis) compliance with the law, conduct a rigorous study of the quality of local implementation, and conduct a study of the consequences of the implementation of the statewide system of educator evaluation. This should include a complete audit of the system three years after full implementation. The audit should examine whether the system improves teaching and learning in the state and effectively supports ongoing educator learning and development. The audit should engage key stakeholders, including teachers, administrators, and parents.
- ❖ A standing committee should be established as a successor to the MCEE. The role of this standing committee should be to:
 - Receive reports from the state regarding compliance with the requirements for educator evaluation, the quality of local implementation, and the consequences of the implementation of educator evaluation.
 - Receive feedback from stakeholders and advocacy groups regarding the statewide evaluation system, the waiver process, and any other related issues.

- Convene on a regular basis to review the feedback and reports and to make recommendations regarding changes to the requirements for educator evaluation. This should include assessing student growth recommendations and their consequences after one year, reviewing the results of the full system audit after three years, and making recommendations for adjustments to the system as warranted by these reviews.
- ❖ The state should gather data on the technical quality of the administrator evaluation tools. Even more so than with the teacher observation tools, there exists little to no empirical evidence about the reliability and validity of the administrator tools and evaluation systems that currently exist. The state will need to be vigilant in collecting and analyzing data to track the qualities of the tools and processes recommended here. As the research base accumulates, it will also be important to revisit these recommendations and recommend improved tools and processes when supported by evidence.
- ❖ A committee should be appointed to study and make recommendations about how to create and build a clear vision of the policies and practices that would be needed for advanced certification to be of benefit to Michigan's schools. Fully developing well-informed recommendations for professional certification was beyond the scope of what the MCEE could accomplish in its brief tenure.

MULTIPLE FUNCTIONS

Here we list and describe the breadth of functions that this capacity-building and infrastructure will need to address.

Information: Statewide, educators and the public will need information about the educator evaluation system. With any large-scale change, there are heightened information needs. Parents and the public will need to understand the system and its implications both for their children and for the state's schools. Teachers and school leaders will need information about the law, the policies in place to implement the law, and their rights and responsibilities. School boards will need information about the system as well, and their responsibilities and/or role in overseeing the processes and learning from the system as it unfolds. Evaluators (principals, as well as others) will need information about training and available resources. Researchers will need information about the state policies, where data are located and how they are shared, and how the system is managed and monitored. Relevant issues here include the creation and upkeep of an educator evaluation system website and the development of materials for informational purposes only, as well as the ongoing development and posting of FAQs for the various relevant users: school board members, teachers, principals, other leaders, superintendents, legislators, and the public.

Observation tools/instruments and their ongoing revision and development: The educator evaluation system will depend on the standardized use of instruments for documentation and evaluation of educators. Some instruments will need to be regularly purchased; some will need to be developed. These instruments will need to be stored and available to LEAs and educators statewide. This also includes training manuals for teachers and evaluators, training manuals for the evaluation of leaders, written instructions about how to follow the policies, training modules for teachers, leaders, and administrators, etc. Part of the question about what will need to be developed depends on what instruments are bought from vendors and what the state and LEAs get as part of contracting with specific vendors. As the educator evaluation system unfolds, new instruments and tools will need to be created to meet needs, and other instruments will need to be adapted and revised.

Student growth assessments and their ongoing revision and development: In order to expand and further standardize the use of student growth measures and VAMs in educator evaluations, the state must continue to invest in the development of assessments that are aligned to state-adopted content standards in all core content areas (i.e., English language arts, mathematics, science, and social studies) and in high-volume non-core content areas where state-adopted content standards exist (e.g., arts, health and physical education, career and technical education, and many high school electives).

Coordination: The educator evaluation system will be necessarily linked to several other statewide systems, most notably the student testing system and any and all data systems related to the educator workforce. Student testing will have to be scheduled in ways that enable using data for teacher effects as well. There will have to be regular articulation and communication with the Michigan Department of Education and the Center for Educational Performance and Information, as well as with the vendors of all instruments that are used, LEAs, the State Board of Education, the Governor's Office and the legislature, state professional educator associations, and teacher unions.

Platforms: Administrators and teachers will need support in managing the information that needs to be entered and maintained for the system. Many states have developed platforms that are used by both teachers and administrators. These platforms include dashboards for administrators that contain comprehensive caseload management tools that provide direct access to all required components of each associated educator's evaluation record; educator dashboards include links to all required evaluation activities with a task list displaying real time completion status. The platforms also allow administrators and teachers to enter information about SLOs, upload relevant artifacts, and enter ratings on evaluation rubrics. The platforms also can combine observation ratings to determine overall teacher ratings and changes in SLOs, as well as allowing users to explore and analyze trends in any data element or elements captured as reportable fields and to compile and track completion of evaluation components (or full evaluations) for individual users or selected groups. (NOTE: Several states have developed platforms that allow for different LEAs to be using different protocols and tools.)

Data storage: The educator evaluation system will put considerable pressure on both the state and LEAs in terms of data storage. Teachers will be observed multiple times. Additional data will be collected through the year-long evaluation process. Records will need to be kept of discussions between principals/evaluators and educators. Data submitted by school leaders will need to be stored. Final ratings based on this information will need to be calculated and stored, and then combined with ratings from student data. Some of these data might be stored at the LEA level, but the state will need to have some "cut" of the information, and both LEAs and the state will need to keep longitudinal records of teachers and schools.

Data analyses: While the portal created by the state for users like administrators and educators will be able to do school-relevant analyses, the state will need to link those data to student test scores. The state will also need to be able to conduct additional analyses to determine statewide trends and to examine how well the system is functioning. Some analyses might be conducted to examine the effects of the educator evaluation system on student learning or to look at the association of teacher preparation or professional development programs with the results of the educator evaluation system. Current state data systems will need to be modified to capture student attendance data, assign proportional responsibility to teachers for their students based on their instructional contact with the students and the nature of their classrooms (e.g., intact, team teaching, special education), and assign subject responsibility. New data systems will need to be developed to allow the storage and analysis of teacher observation data, student growth and VAM data, and administrator evaluation data.

Training: Putting the educator evaluation system in place will require statewide capacity building. Educators need to understand the law, and how the policies enacted meet the letter of the law. They will need to understand their rights and responsibilities, and how to interact with and use all relevant educator evaluation system tools and instruments. Disciplined use of the instruments is required to maintain reliability and validity and to ensure equitable and professionally responsible treatment of all participants. Retraining will also be needed, as experience has shown that there is a degradation of reliable tool use by evaluators over time. Some training might be offered through vendors, and other training will need to be done by the state. Some other states have developed web-based training modules that can be completed by large numbers of users at times convenient to them. Other states have had regional training sessions offered. Some examples of the types of expertise that educators must develop include:

- Using observation tools and protocols;
- Using locally developed measures of student growth;

- Understanding state provided VAM data; and
- Developing and using student learning objectives.

Technical support: Questions of all sorts will inevitably arise. These might involve issues of procedure and policy, or technical problems with particular instruments, or questions about SLOs or VAMs. Some questions might best be answered by the relevant vendors, but questions from the field are usually “bundled,” including technical issues and policy/procedure issues and personnel issues. FAQs can be developed and distributed/posted as part of the “information” function, but ongoing questions will need to be answered in a timely fashion by knowledgeable experts who might be diffusely spread; this, too, will require coordination.

VAMs: Considerable work within the existing system will need to be done to tie student assessments to educator evaluation. This will include coordinating with the student assessment offices, aligning policies for testing students with timing that will allow for judgments concerning specific teachers, checking rosters of which students were in which teachers’ classes and how long, handling the grade levels and content domains that are not tested, and coordinating with LEA-level decisions. Here especially, responsibility will be shared across areas and it will be crucial to have people in place that can coordinate efforts, work in responsive ways needed for such interdependent work, and nimbly face the challenges that will arise.

Controlling the quality of assessment administration: As has already been seen in several states, high-stakes use of assessments can provide incentives for inappropriate behavior in the preparation for and administration of assessments. The MCEE recommends that the state develop a model security policy delineating appropriate and inappropriate activities in preparing for and administering state-provided assessments. Likewise, LEAs should adopt a similar policy for any and all assessments that enter into educator evaluations. In addition:

- The state should develop a clear policy for voiding results on state-provided assessments in the event of violations of security.
- The state should be granted authority to monitor security compliance on state-provided assessments.
- The state should be granted authority to monitor local compliance with local security policies.
- LEAs should be provided flexibility in determining individual consequences for violations of security.

Approving waivers, monitoring implementation, and compliance: Some authority will need to receive, review, and approve/disapprove waivers for educator evaluations, as well as to follow up on annual work in LEAs that receive waivers. As with all mandated accountability systems, there will also need to be staff monitoring compliance, which might include random audits of particular LEAs to assess compliance. Any school/LEA out of compliance or struggling with strong implementation will need support.

Managing low quality: As a consequence of these systems, some educators will be identified as weak; others—having been provided opportunities to improve—will be dismissed. LEAs will bear the burden of identifying these teachers and administrators and putting explicit improvement plans in place with hard deadlines for change. The state will need to ensure that LEAs act on identified weaknesses, and that ineffective educators are dismissed in a timely and just fashion.

Complaints and challenges: Educators who wish to object to their rating or to raise complaints about the integrity of the process in their schools/LEAs will need a process by which complaints are registered, heard, and adjudicated. The MCEE recommends that this be a two-stage process, where challenges are first made at the LEA level. If the objection is not satisfactorily resolved, then an educator may take the complaint to the state level. The state will need a process for reviewing and settling such challenges.

And, finally, while these functions are described at the state level, there will be capacity and infrastructure implications at school and LEA levels as well that need to be addressed. That is, there will need to be

LEA expertise in handling the data, communication, articulation, and policy demands implicated in a statewide educator evaluation system. Though some of that expertise can be developed in training around instruments, a great deal of it will need to be developed as the system is implemented. LEA expertise will be especially dependent on strong communication and information support from the state.

KEY JURISDICTION RECOMMENDATIONS

Given the sprawling nature of the interconnections—student testing and teacher/administrator issues have not historically been located in the same offices—it is unlikely that all of these functions will be done by one office. While there is no single right answer to the question of how this system should look, the MCEE does recommend a set of design principles that should be used in expanding the state system to accommodate these new demands. It is imperative that the infrastructure be attended to carefully, as the success of putting a rigorous educator evaluation system into place depends on building an infrastructure that can enable and sustain it.

- ❖ Setting up and supporting this system will take considerable resources. The state will need to set aside both start-up and long-term funds to support the human and material resources necessary.
- ❖ It is essential that a user-friendly interface be developed for interactions with LEAs, educators, and the public. The learning demands of putting this system in place are high, and successful implementation will require knowledgeable, responsive staff with high quality and clear information.
- ❖ It is essential that the infrastructure, staffing, and resources are developed to encourage improvement, not compliance. This means keeping forms clear and short, for example, and focusing evaluations on a small set of instruments that produce the most crucial and trustworthy information. Lengthy paperwork put on the shoulders of educators will not improve practice. Streamlined tools that enable strong feedback and support are necessary.
- ❖ The expertise to put this system in place is also considerable. It is essential that various functions listed above are done with and where people have the most relevant expertise.
- ❖ The infrastructure necessary to support educator evaluation systems will transcend single departments in state offices. This means that very strong “connective tissue” will need to be built across units. It is imperative to develop clear lines of authority and responsibility, to have policies and practices that encourage regular communication, and to hire staff that are talented in proactive and positive team building and collaboration. The public face of this system must be user friendly, change oriented, able to speak to multiple stakeholders, and responsive and focused on positive change in the name of student learning.
- ❖ There will be a great deal learned in the first five years of establishing this system. This might mean that the infrastructure evolves over time. It is essential that the state collect relevant information and data to keep track of what we learn. Some of these data will contribute to changes in tools; some will contribute to research on the technical merits of instruments and the system more generally; some will inform us about how best to educate the public, to train evaluators, to support teachers and administrators who need improvement, and to work with teacher preparation, induction, and professional development programs to support learning aligned with this system.

An important part of the new educator evaluation system will be establishing a standing committee to succeed the MCEE that will oversee, first, the substantial start up that this will involve, and, later, the quality of the system and ensure that, as new research and technology is available for rigorous teacher and administrator evaluation, the state policies and practices are changed. This new standing committee will also act as a mediator to resolve issues that arise across units as they learn to coordinate these new activities and functions and can also be the mechanism to maintain the ongoing relationship with the executive branch.

Appendix A: Council Members' Biographies

Deborah Loewenberg Ball is the William H. Payne Collegiate Professor in education at the University of Michigan, and an Arthur F. Thurnau Professor. She currently serves as dean of the School of Education and as director of a new organization called TeachingWorks. She taught elementary school for more than 15 years, and continues to teach mathematics to elementary students every summer. Ball's research focuses on the practice of mathematics instruction, and on the improvement of teacher training and development. She is an expert on teacher education, with a particular interest in how professional training and experience combine to equip beginning teachers with the skills and knowledge needed for responsible practice. Ball has served on several national and international commissions and panels focused on policy initiatives and the improvement of education, including the National Mathematics Advisory Panel and the National Science Board.

Jennifer Hammond is the principal of Grand Blanc High School. She previously served as a teacher and administrator at schools in Troy, Hamtramck, and also in Houston, Texas. Hammond earned a bachelor's degree and certificate in secondary teaching from Michigan State University, a master's degree in mathematics education from Wayne State University, an educational specialist degree in school administration from Oakland University, and a doctorate of philosophy in educational leadership from Oakland University.

Joseph Martineau is the deputy superintendent for accountability services in the Michigan Department of Education. He has served in the Michigan Department of Education as a psychometrician, manager of large-scale assessment programs, and director of state testing and accountability. He also serves as a member of the board of the National Council on Measurement in Education, and on the executive committee of the Smarter Balanced Assessment Consortium. Martineau earned a bachelor's degree in linguistics and a master's degree in instructional design from Brigham Young University and a doctorate in measurement and quantitative methods from Michigan State University. Martineau serves on the council as a non-voting member as the designee of the state superintendent of public instruction.

Mark Reckase is a professor in the measurement and quantitative methods program within the Counseling, Educational Psychology, and Special Education Department of the College of Education at Michigan State University. He worked for 17 years at ACT Inc., a college admission testing company and was a faculty member at the University of Missouri-Columbia. Reckase also served as the vice president of the American Educational Research Association and the president of the National Council of Measurement in Education. He earned a bachelor's degree in psychology from the University of Illinois, and a master's degree and doctorate in psychology from Syracuse University.

Nicholas Sheltrown is director of measurement, research, and accountability at National Heritage Academies in Grand Rapids. He manages the measurement and research initiatives for a network of 71 charter schools with over 40,000 students. Sheltrown previously served as director of research and measurement at Grand Valley State University, the technology director at Byron Center Public Schools, and vice president of professional development at ST Concepts Inc. in Byron Center. He earned a bachelor's degree in mathematics from Cornerstone University, and a master's degree in curriculum and teaching and a doctorate from Michigan State University.

David Vensel is the principal of Jefferson High School in Monroe. He previously served as a teacher and assistant high school principal at Airport High School in Carleton. He earned a bachelor's degree in sociology from Eastern Michigan University and master's degree in American history and secondary education from the University of Toledo.

Appendix B: Annotated Agendas

Governor's Council on Educator Effectiveness
Wednesday, December 7, 2011 • 2:00 – 5:00 p.m.
Lansing, Michigan
CLOSED SESSION: George W. Romney Building • 111 S. Capitol Ave.

AGENDA

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltroun, Dave Vensel

2:00 – 2:30 Welcome, introductions, and preview of charge

Council members introduce themselves and share brief details about the expertise they bring to the work of the Governor's Council.
Deborah Loewenberg Ball previews the work of the coming months.

Notes: The word "tool" does not necessarily mean that we will suggest one tool, but that we will develop principles that guide the legislature. The GCEE is contributing to the infrastructure for training, development, and evaluation of teachers. A checklist is not sufficient to measure effectiveness. The GCEE agrees that it is very important to build consensus around this work.

2:30 – 3:00 Framing: The challenges of teacher evaluation

What are the greatest challenges in developing principles for a teacher evaluation system?

Notes: The legislation makes this a political charge. Perhaps the council can encourage less partisan features of the legislation.

3:00 – 3:20 Review council curriculum and procedures and finalize meeting schedule

3:20 – 3:30 Move to Capitol Building

PUBLIC SESSION: Capitol Building • 100 N. Capitol Ave. • Rooms 402 and 403

3:30 – 3:40 Review of charge and introduction of council members

Deborah Loewenberg Ball reads the official charge of the Governor's Council.
Council members introduce themselves to invited speakers and guests.

3:40 – 4:40 Prepared remarks from invited speakers

Representatives from key groups who have a stake in the work of the council make brief prepared statements. They include:

- Phil Pavlov, Senator, 25th District; Chair, Senate Education Committee
- Paul Scott, former Representative, 51st District
- Debbie Squires, Associate Director, Michigan Elementary and Middle School Principals Association
- James N. Goenner, President & CEO, National Charter Schools Institute
- Dan Quisenberry, President, Michigan Association of Public School Academies
- Brad Biladeau, Associate Executive for Government Relations, Michigan Association of School Administrators
- Jim Ballard, Executive Director, Michigan Association of Secondary School Principals (or alternate)
- Amber Arellano, Executive Director, The Education Trust-Midwest

- Chad Aldis, State Director, StudentsFirst
- Dan Varner, CEO, Excellent Schools Detroit
- Louise Somalski, Legislative Coordinator, AFT Michigan
- Art Przybylowicz, Associate Executive Director and General Counsel, Michigan Education Association

Notes: Speakers suggested creating a fair, transparent, valid, and reliable system. Empower principals to become instructional leaders, and use evaluation as a development tool. Study what other states have implemented and learn from them.

4:40 – 5:00 Public remarks

Open the floor for brief remarks from others in attendance.

Next meeting Wednesday, December 14, 2011

2:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Governor's Council on Educator Effectiveness
Wednesday, January 11, 2012 • 2:00 – 5:00 p.m.
University of Michigan School of Education**

Dean's Conference Room (room 1211) • 610 East University Avenue, Room 1211 • Ann Arbor

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

2:00 – 2:20 Opening to meeting and updates

Changes in GCEE structure; funding; consultants; communication protocols

2:20 – 3:20 Purposes of evaluations

Why is it important for states and/or school districts to develop evaluation systems for their educators and administrators? What are the key purposes for such evaluations? What professional standards (technical, legal, and ethical) should guide the use of evaluations?

Discussion led by Brian Rowan, Burke A. Hinsdale Collegiate Professor at University of Michigan School of Education

To review in advance: "Measuring What Matters" (December 2010/January 2011 issue of *Kappan*) and "Evaluating Teacher Effectiveness: Where do we go from here?" (National Comprehensive Center for Teacher Quality's May 2011 presentation to Learning First Alliance)

Notes: The key purpose of evaluation systems is to improve teaching and learning. The council should outline the things that need to be in place in order to implement evaluations validly. Start with standards, and use these to select a tool.

3:20 – 3:45 Review of the legislation

The GCEE was established as part of Michigan's teacher tenure reform efforts (PA 102 of 2011). What does the legislation require the GCEE to include in its recommendations? What does a close reading of PA 102 and the bill analysis teach us about the intent of the legislation?

Discussion led by Deborah Loewenberg Ball

To review in advance: PA 102 and analysis of the legislation

Notes: Start with the definition of effective teachers and tailor this definition for different instruments. Find out what domains are being measured in other states.

3:45 – 4:45 Other states' efforts

A number of states have already developed evaluation systems. What is typically assessed by these systems? Do any states provide a model for us to follow as we develop our recommendations?

Discussion led by Suzanne Wilson, University Distinguished Professor, chair of the Department of Teacher Education, and director of the College of Education's Center for the Scholarship of Teaching at Michigan State University

To review in advance: National Comprehensive Center for Teacher Quality's comparison of teacher evaluation policies for Rhode Island, New York, and North Carolina (To compare other states, visit <http://resource.tqsource.org/stateevaldb/>)

Notes: North Carolina is a high capacity state with partnerships with research universities. We will need to determine Michigan's capacity. Rhode Island is a good model and clearly lays out its methodology. Rhode Island uses three tools for observations. New York has five observation tools that districts can use.

4:45 – 5:00

Our charge

The GCEE is charged with identifying recommendations for all of the following:

- 1) A student growth and assessment tool.
- 2) A state evaluation tool for teachers.
- 3) A state evaluation tool for school administrators.
- 4) Recommended changes to be made in the requirements for a professional teaching certificate.
- 5) A process for evaluating and approving local evaluation tools for teachers and administrators that are consistent with the state evaluation tool for teachers and administrators and the act.

What will count as a recommendation? What principles should guide our work?

Discussion led by Deborah Loewenberg Ball

Notes: The GCEE needs to make sure there is empirical evidence that the instrument is valid. This poses a challenge with both choosing and building our own.

Next meeting

Wednesday, January 18, 2012

2:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

Focus: Key types of teacher evaluation tools and/or systems

**Governor's Council on Educator Effectiveness
Wednesday, January 18, 2012 • 2:00 – 4:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Mark Reckase, Joseph Martineau, Nick Sheltrown, Dave Vensel

AGENDA

2:00 – 2:10 Opening to meeting and updates

2:10 – 2:40 Walkthrough tool used at Monroe Public Schools

Presentation by Julie Everly, assistant superintendent for elementary education, and **Ryan McLeod**, assistant superintendent for secondary education, Monroe Public Schools

Notes: Monroe Public Schools has an iPad walkthrough tool for observations. Tools prompt district level “look fors” and allow others to be added at the school level. MPS will be drafting a rubric based walk-through tool next in order to get away from the yes/no model. The district asks principals to do ten walk-throughs each week. This model allows immediate feedback for teachers.

2:40 – 3:15 Two rubrics: Danielson and Marshall

Compare and contrast the two rubrics. Where are the overlaps? What's missing? How do they align with the chart Brian presented at the January 11 meeting?

Discussion led by Brian Rowan, Burke A. Hinsdale Collegiate Professor at University of Michigan School of Education

To review in advance (IN DROPBOX):

- Charlotte Danielson's Framework for Teaching (2011 Revised Edition)
- Kim Marshall's Teacher Evaluation Rubrics (revised September 4, 2010)

Notes: Marshall doesn't use evidence, only judgment. Danielson has a statement of standards and is well developed and elaborated. According to the MET study, observation tools should define expectations for teachers, ensure observer accuracy, ensure reliability of results, and determine alignment of outcomes. Ensuring accuracy of observers is a huge challenge, but MET recommends that observers be trained and certified.

3:15 – 4:00 Three models: North Carolina, Rhode Island, and Washington, D.C.:

Compare and contrast the three models. Where are the overlaps? What's missing? How well do they address some of the concerns placed in the “parking lot” at the January 11 meeting (e.g., reliability of data, transparency of process, validity of instrument, application to untested grades and subjects)?

Discussion led by Suzanne Wilson, University Distinguished Professor, chair of the Department of Teacher Education, and director of the College of Education's Center for the Scholarship of Teaching at Michigan State University

To review in advance (IN DROPBOX):

- North Carolina Teacher Evaluation Process

- The Rhode Island Model: Guide to Evaluating Building Administrators and Teachers (2011-2012)
- IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel (Group 1: General Education Teachers with Individual Value-Added Student Achievement Data)
- IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel (Group 2: General Education Teachers without Individual Value-Added Student Achievement Data)

Notes: Washington, D.C. model is concrete and describes behaviors and examples in depth. North Carolina looks like National Board and focuses on teachers as leaders. Rhode Island seems oriented toward developing over time and learning.

Next meeting **Wednesday, February 8, 2012**
2:00 – 5:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Governor's Council on Educator Effectiveness
Wednesday, February 8, 2012 • 2:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltrown, Dave Vensel

AGENDA

2:00 – 2:10 Opening to meeting and updates

2:10 – 3:10 Big issues

What big issues or questions need to be addressed before we can make any recommendations about principles or tools for evaluation? We have already begun building a “parking lot” for these that includes things like access to data, transparency, and validating evaluation instruments. What other big categories do we need to consider? What fundamental questions concern you most about this work?

Notes: Council members' questions include: How do we deal with the differences in context, level, and subject matter? Are we developing our own tool, or are we looking for a tool or approach that is already developed? To whom do our recommendations apply? How do we communicate with the legislature, teacher organizations, and others? Regarding the student growth tool, what is the metric? Some next steps are to create a vision statement, continue to look into what other states have done, and continue to research existing tools.

3:10 – 3:40 Guiding principles

At our first meeting, I said that any recommendation that we make needs to be valid, fair, useful, and feasible. Are there other principles that should guide our work?

3:40 – 4:10 Learning from experts

What two or three things are you most needing to learn about from consultants or each other to do this work responsibly? Do you have suggestions for experts we could bring in to guide some of that learning?

4:10 – 5:00 Advisory Committee

What role do you envision for the soon-to-be-appointed advisory committee of teachers, administrators, and parents?

Notes: The Advisory Committee can identify the concerns and expectations that they have; this could give the GCEE insight into what others are worrying about and hoping for. The Advisory Committee could develop a plan to learn about what a subset of districts is doing now, and use that to inform a list of components that they believe should be included in an evaluation system. The GCEE needs to learn how best to work with the Advisory Committee.

Next meeting Monday, February 13, 2012

2:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Governor's Council on Educator Effectiveness
Thursday, February 13, 2012 • 2:00 p.m. – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

2:00 – 2:10 Opening to meeting and updates

2:10 – 5:00 VAM team and observation tool teams conduct small group work

**Governor's Council on Educator Effectiveness
Thursday, February 16, 2012 • 9:00 a.m. – 12:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members Present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltroun, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates

9:10 – 9:20 Timeline for deliverables and resulting political tensions

Discuss concerns we have heard from some legislators regarding a revised timeline that allows us to complete our work by the end of the calendar year. Consider strategies for addressing concerns.

Notes: The council decided that the timeline for deliverables needs to be extended in order to make responsible recommendations. Deborah could make this proposal at her March 1 meeting.

9:20 – 9:45 Communication strategies and guidelines

Review and comment on vision statement drafted by Jenny and Dave.
Discuss key talking points, protocols for media requests and other official communications on behalf of the council, meetings with key stakeholder groups, etc.

To review in advance (in Dropbox): Draft vision statement

9:45 – 10:00 Timing of statewide student tests

Discuss the benefits and drawbacks of administering statewide tests (e.g., MEAP, ACT, MME) in the spring

Notes: MEAP will move online and to the spring in 2014-15. The state could provide some funding to do benchmark/periodic assessments in non-tested grades and subjects. MCEE will continue to consider assessment timelines and their alignment with evaluation recommendations.

10:00 – 11:15 Work in small groups

11:15 – 12:00 Presentation by David Hecker, president, Education Alliance of Michigan

Notes: Districts will need a lot of support to use valid and reliable assessments in all content areas. The council must have a mobility standard; many classrooms, especially in urban districts, change composition over the course of the year. The GCEE should consider using peer reviews, portfolios, and self-assessments.

Next meeting Tuesday, February 21, 2012

2:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

Note: Dan McCaffrey, PNC Chair in Policy Analysis and senior statistician at RAND Corporation, will be presenting the Frank B. Womer Lecture at the School of Education from 12:00 to 1:30 p.m. Governor's Council members are invited to attend.

**Governor's Council on Educator Effectiveness
Tuesday, February 21, 2012 • 2:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

- 12:00 – 1:30 Daniel F. McCaffrey presents the 2012 Frank B. Womer Lecture in Measurement and Assessment**
Dan McCaffrey, PNC Chair in Policy Analysis and senior statistician at RAND Corporation will present his talk, "Can Paying Teachers for Performance Improve Student Achievement? Results from Three Random Assignment Evaluations." All GCEE members are invited. His talk will take place in the Prechter Laboratory (room 2202) in the School of Education. A light lunch will be served.

For more information, see:

http://soe.umich.edu/news_events/events/detail/womer_lecture_daniel_mccaffrey/

NOTE: Dan McCaffrey will join us for the GCEE meeting after his talk and answer questions we have about value-added modeling and other student growth models. Nick and Joseph prepared some questions in advance, which are included on this agenda. Please feel free to bring your own questions to the meeting.

- 2:00 – 2:10 Opening to meeting and updates**
Proposal to hire GCEE project manager

- 2:10 – 3:10 Introduction to value-added modeling (VAM)**
- What technically qualifies a model to be value-added? What minimum characteristics must a model have to be considered a value-added model?
 - What is the simplest value-added model that could be used responsibly in educator evaluation?
 - What factors should we consider when selecting a value-added model?
 - What potential benefits does VAM present to a teacher evaluation system?
 - What are the potential pitfalls?

To review in advance (in Dropbox): Daniel Koretz's 2008 *American Educator* article, "A Measured Approach"

Notes: There is no universal definition of VAM, but there are components that everyone agrees are a part of VAM (e.g. this year's scores regressed against last year's and the year's before with a consideration for demographics). Dan suggests: regress the current year score on some set of prior year scores, account for error in prior scores, add aggregated scores at classroom level to control for peers.

- 3:10 – 4:10 Using VAM to evaluate and improve instruction**
- How would you suggest value-added data be incorporated in a teacher evaluation system?
 - What advice would you give practicing educators who must incorporate VAM in an overall evaluation?
 - How would you recommend using VAM to provide feedback to teachers to help them improve instruction?

- Most VAMs compare teachers against the average teacher effect, but how do you know if the average teacher is effective?
- How much does choice in what VAM model you select influence things like teacher ranking and evaluation?
- What do you think about the role of “growth toward a standard” models?

4:10 – 5:00 VAM data integrity and reliability

- What is a simple design that doesn't require additional data collection to test the effect of a VAM-based system in the state (e.g., interrupted time series design)?
- What are the biggest data quality issues that you have encountered that compromise VAM?
- In the 2003 report, *Evaluating Value-Added Models for Teacher Accountability*, you wrote, “The research base is currently insufficient to support the use of VAM for high-stakes decisions.” Is this still true in your opinion?

Notes: The GCEE should focus on error where stakes are the highest. Some other factors to consider include putting in peer effects, accounting for students with multiple teachers, precision, and statistical bias.

Next meeting Monday, February 27, 2012, 2:00 – 5:00 p.m.

Grand Valley State University, Eberhard Building, room 215

301 Fulton St. W, Grand Rapids, MI

(see <http://www.gvsu.edu/meetatgvsu/eberhard-parking-directions-and-map-12.htm> for a map and parking information)

**Governor's Council on Educator Effectiveness
Monday, February 27, 2012 • 9:00 a.m. – 12:00 p.m.
Grand Valley State University • Eberhard Building, room 215 • 301 Fulton St. W • Grand Rapids**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and welcome remarks

9:10 – 9:40 Use of teaching evaluations and student achievement scores to improve instruction

Presentation by Tom Livezey, superintendent, and **Jason Kennedy**, principal, Oakridge Public Schools, Muskegon, MI

9:40 – 10:30 Observation tools and other modes for measuring the effectiveness of instruction

Discussion led by Suzanne Wilson, University Distinguished Professor, chair of the Department of Teacher Education, and director of the College of Education's Center for the Scholarship of Teaching at Michigan State University

Notes: Suzanne Wilson guided the group in a discussion of observation tools. Council members developed a list of questions for observation tool developers. Answers to these will assist the GCEE in determining which tools might best fit Michigan districts and schools.

10:30 – 11:30 Student growth and assessment tools

Discussion led by Deborah Loewenberg Ball, chair, Governor's Council on Educator Effectiveness; William H. Payne Collegiate Professor of Education, Arthur F. Thurnau Professor, and dean, University of Michigan School of Education

Notes: Review Dan McCaffrey's talk.

11:30 – 12:00 Public comment session

Notes: Suggestions from public attendees included looking at student growth percentile model as an interim student growth option, examining the state's professional development opportunities, using multiple observers and student/parent surveys.

**Next meeting Thursday, March 1, 2012
9:00 a.m. – 12:00 p.m.
Lansing, MI (exact location to be determined soon)**

**Governor's Council on Educator Effectiveness
Thursday, March 1, 2012 • 9:00 a.m. – 12:00 p.m.
Capitol Building, room 424 • 100 North Capitol Avenue • Lansing, Michigan**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltrown, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates
Debriefing Monday's meeting in Grand Rapids
Updates

9:10 – 9:40 Systematized evaluation: National Heritage Academies

Presentation by Max Hunsicker, senior director of coaching and learning at National Heritage Academies

Notes: According to Mr. Hunsicker, National Heritage Academies' evaluation is intentional, supportive, and measured. The goal of this system is to have high-quality teachers in every classroom. The system focuses on components of teaching that have the greatest impact on student achievement. This system is built around meaningful dialogue and professional development.

9:40 – 10:10 Update on meeting with legislators
Notes from conversation with Senator Phil Pavlov, chair of the Senate Education Committee
Highlights from meeting with key legislators

10:10 – 11:00 Outstanding questions and next steps
Review questions surfaced at Monday's meeting
Determine assignments and next steps

To review in advance: Grids of questions about observation tools and student growth models (in Dropbox in folders "Observation tool questions" and "Student growth questions")

Notes: Council members reviewed this question grid and determined assignments for future work. The primary focus for upcoming weeks will be on observation tools.

11:00 – 12:00 Student growth and value-added models
Review notes from conversation with Dan McCaffrey
Begin building framework for building recommendations for feasible and useful student growth assessments

**Next meeting Wednesday, March 7, 2012
2:00 – 5:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)**

**Governor's Council on Educator Effectiveness
Wednesday, March 7, 2012 • 2:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room (room 1211) • 610 East University Avenue • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

2:00 – 2:10 Opening to meeting and updates

Debriefing last week's meetings in Grand Rapids and Lansing
Updates

2:10 – 2:40 Update on meeting with legislators

Notes from conversation with Senator Phil Pavlov, chair of the Senate Education Committee
Highlights from meeting with key legislators
Next steps

2:40 – 3:00 Michigan assessment timeline through 2015-16

Discuss the state's assessment timeline and its impact on student growth models

To review in advance: Assessment timeline (in Dropbox folder "Relevant MDE policies")

3:00 – 4:00 Washington perspective

In 2007, the Center for Educational Leadership (CEL) at University of Washington College of Education released its instructional framework, the 5 Dimensions of Teaching and Learning (5D). According to CEL's website (www.k-12leadership.org):

We will have a Skype conversation with **Steve Fink**, executive director at CEL, **Sandy Austin**, project director at CEL, and **Edie Holcomb**, program facilitator at Washington's Teacher & Principal Evaluation Pilot (TPEP).

To review in advance: Materials from University of Washington (in Dropbox folder "University of Washington")

Notes: Washington is using three instructional frameworks, but 5D reflects the overall scope of Danielson and Marzano. In Washington, these frameworks will be used with all instructional personnel. Each of the providers (Danielson, Marzano, and 5D) will provide training. The instrument is not as important as the training to use the framework well. These presenters believe that observers do not judge a classroom, but watch and catalogue.

4:00 – 5:00 Work in small groups

Next meeting Friday, March 16, 2012

2:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Governor's Council on Educator Effectiveness
Friday, March 16, 2012 • 2:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room (room 1211) • 610 East University Avenue • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown

AGENDA

2:00 – 2:10 Opening to meeting and updates

Introduce Cori Mehan
Updates

2:10 – 2:30 Michigan assessment timeline through 2015-16

Discuss the state's assessment timeline and its impact on student growth models

To review in advance: Assessment timeline (in Dropbox folder "Relevant MDE policies")

Notes: Council members studied the testing timeline and asked Joseph questions about implementation and feasibility. It was noted that this timeline will help council members as they make future recommendations regarding student growth.

2:30 – 3:00 Review of Michigan's current data availability and challenges

Nick Sheltrown and Joseph Martineau will present and lead a discussion

Notes: Nick and Joseph explained roster checking, which would allow for districts to match students more accurately with teachers and glean a more accurate measurement for each teacher's percentage of instructional responsibility. The council discussed the difficulty of applying such a tool to PE and art teachers, but decided to consider roster verification tools as they continue to make recommendations.

3:00 – 4:00 "Teaching capacity" growth model

Mark Reckase and Joseph Martineau will present an alternative growth model that they are developing to measure "teaching capacity"

Notes: Mark and Joseph presented their growth model and answered questions. They explained that this model would allow districts to consider and account for students' backgrounds and other external factors when evaluating student growth. Each student would receive a challenge index. One outstanding concern was that this model might favor teachers working with disadvantaged student populations.

4:00 – 5:00 Colorado perspective

Colorado's State Council for Educator Effectiveness submitted its report and recommendations to the State Board of Education on April 13, 2011. We will have a Skype conversation with **Lorrie Shepard**, member of the council and dean at the University of Colorado at Boulder's School of Education. She will offer information about their council's efforts, the process they used to arrive at their recommendations, and key lessons learned.

To review in advance: Attached summary of Colorado's State Council for Educator Effectiveness Report and Recommendations (Full report is in Dropbox folder "Sample reports of Ed Evaluation Committees")

Notes: Lorrie Shepard explained the educator evaluation process in Colorado, including their timeline, matrix approach, pilot, and choosing an observation tool.

Next meeting **Wednesday, March 28, 2012**
10:00 a.m. – 1:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

Michigan Council for Educator Effectiveness
Wednesday, March 28, 2012 • 10:00 a.m. – 1:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown (via telephone), Dave Vensel (via telephone)

AGENDA

10:00 – 10:45 Opening to meeting and updates

Executive Order changing our name; Plans for open meeting in Detroit on April 2;
Discussion of ways to engage the Advisory Council before the April 30 deadline;
Summer meeting dates

Notes: The governor signed an executive order that changed the council's name to the Michigan Council for Educator Effectiveness. Our logo must be changed, as well as other documents. The Detroit meeting will take place at the Skillman Foundation. Cori will send out directions and parking information. The MCEE will ask the Advisory Council for their input on key challenges.

10:45 – 11:00 Vision statement approval

To review in advance (In Dropbox folder "Vision statements"): Vision statement revisions document

Notes: Council members edited and approved the vision statement, which will guide the council's future work and recommendations.

11:00 – 11:30 Combined performance measures

Nick Sheltrown will review how five states combine their performance measures.

Notes: Nick provided information on combining performance data. In the discussion afterward, the council generally preferred the rubric approach, not the formula approach. The council also agreed on the need to be able to indicate the probability that a teacher will fall into any given box in the rubric. For future thinking, could this approach set Michigan apart from other states?

11:30 – 1:00 Review of observation tool conversations and findings

Jenny Hammond, Dave Vensel, and Suzanne Wilson will review the observation protocols and frameworks that they have examined and discuss findings, thoughts, and questions.

To review in advance:

- Memo concerning observation protocols and related materials/processes (to be emailed later on March 27)
- Charlotte Danielson's Framework for Teaching (In Dropbox folder "Meeting agendas and materials")
- Robert J. Marzano's An Observational Protocol Based on "The Art and Science of Teaching" (In Dropbox folder "Meeting agendas and materials")
- University of Washington's 5D+ Teacher Evaluation Rubric (In Dropbox folder "Meeting agendas and materials")

Note: If you received a binder that contains these observation tools, please bring it with you to the meeting.

Notes: Jenny, Dave, and Suzanne met with representatives from observation tool organizations to learn more about the specifics of each tool. Council members discussed observation tool ideas, concerns, and questions regarding feasibility, reliability, validity, cost, and other aspects of each system.

Next meeting **Monday, April 2, 2012**
12:00 – 3:00 p.m.
The Skillman Foundation (100 Talon Centre Dr., Suite 100, Detroit)

**Michigan Council for Educator Effectiveness
Monday, April 2, 2012 • 12:00 p.m. – 3:00 p.m.
The Skillman Foundation • Grantees' Room • 100 Talon Centre Dr., Suite 100 • Detroit**

Council members present: Deborah Ball, Dave Vensel

AGENDA

12:00 – 12:15 Opening to meeting and welcome remarks

12:15 – 1:00 Updates on the MCEE's work

Discussion led by Deborah Loewenberg Ball, chair, Michigan Council for Educator Effectiveness; William H. Payne Collegiate Professor of Education, Arthur F. Thurnau Professor, and dean, University of Michigan School of Education

Notes: The MCEE has looked at particular observation frames and protocols like Danielson. Council members have looked at other states in order to learn what systems exist. The council is working to develop a system that is fair, transparent, and feasible, and will contribute to educational improvement.

1:00 – 1:30 Learning about the Washington's evaluation pilot

We will have a phone conversation with **Michaela Miller**, Washington's teacher–principal evaluation project manager, to discuss Washington's pilot program. Michaela will discuss timeline, training, cost, feedback from educators, and other lessons learned regarding Washington's educator evaluation pilot.

Notes: Washington has plans to phase in their system; there are nine school districts in the pilot this year and there will be 65 school districts in 2012-2013. Washington was able to train all teachers in pilot schools on the observation tools, but it was expensive. Michaela suggests that the MCEE focus on connecting teachers and principals in pilot districts, use frameworks that already exist, work with teachers to set goals, and listen to feedback from teachers.

1:30 – 2:00 Piloting evaluation systems

- What are the benefits of a pilot year?
- What systems or policies need to be in place for a pilot to be effective?
- How might districts apply to be a part of a pilot year?

Discussion led by Cori Mehan, project manager for the Michigan Council for Educator Effectiveness.

Notes: After examining other states, Cori shared some findings. Selecting varying sizes of pilot districts can help to understand more potential challenges. The cohort of pilot districts should be relatively small so that the state can analyze the evaluation systems' effectiveness in each school. In many cases, student growth measures are not piloted in the first year.

2:00 – 3:00 Public comment session

Notes: Create more transparency with the public. Avoid “gotcha” checklist evaluation. Evaluation system needs to be about professional growth. Pilots are important for buy-in; pilots also help to ensure that a system works before asking more districts to take part.

Next meeting Monday, April 12, 2012
8:00 – 11:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Thursday, April 12, 2012 • 8:00 – 11:00 a.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Nick Sheltrown, Dave Vensel

AGENDA

8:00 – 8:10 Opening to meeting and updates

8:10 – 8:20 Overview of the April 2 meeting in Detroit

Notes: During the Detroit meeting, the council heard from Michaela Miller, who explained the pilot process in Detroit. She answered questions about piloting districts, feedback from educators, and the lessons they learned. Then, the council heard from public attendees including teachers, district leaders, and members of advocacy groups.

8:20 – 10:45 Plans and considerations for the interim progress report

What does the Council want to include in the upcoming report? What recommendations can we make? What can we say regarding the recommendations we are not yet prepared to make? What should be our messaging strategy around this report?

Discussion led by Deborah Loewenberg Ball

Notes: The council agreed that the upcoming interim progress report should describe the council's work, and should include consultants, agendas, and lessons that the council has learned. Sections of the report will include process, observation tool, student growth, timeline, and pilot recommendations. Council members agreed to work on sections of the report, and the draft will be available to view before the next meeting.

10:45 – 11:00 Sharing Social Security Numbers with Jessica Menold

Jessica Menold, finance specialist in the Executive Office of Governor Snyder, is working to reimburse council members for mileage and other expenditures. She needs each council member's social security number, and will be speaking with us via telephone to procure these.

Next meeting Thursday, April 19, 2012

1:00 – 4:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Thursday, April 19, 2012 • 1:00 – 4:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

1:00 – 2:00 Danielson Framework and Teachscape presentation

Charlotte Danielson is the creator of the Danielson Framework and has served as a consultant to hundreds of districts, universities, intermediate agencies, state departments of education, and national ministries and departments of education. She will be speaking with us about her observation framework and the policies and practices that support its implementation.

Also visiting is **Mark Atkinson**, the founder and CEO of Teachscape, an organization that “combines software tools for classroom observation and evaluation, online learning content based on authentic teaching practice, and professional services for support in structuring professional development and implementing school turnaround.” Mr. Atkinson has worked closely with Ms. Danielson to develop an online training, practice, and assessment system for observers to ensure that they can make accurate and consistent judgments based on evidence.

To review in advance (In Dropbox):

- Charlotte Danielson's Framework for Teaching (2011 Revised Edition)
- Questions for Charlotte Danielson

Notes: Ms. Danielson and Mr. Atkinson answered questions regarding the training, cost, and overall implementation for the Framework for Teaching. Mr. Atkinson briefly showed portions of the online training portal. He will give council members access to this portal so that they can review its features and sessions.

2:00 – 4:00 Reading, editing, and continuing to write the interim progress report

Discussion led by Deborah Loewenberg Ball

Notes: The council read and reviewed the interim progress report draft, and Chair Ball made notes throughout the document. Council members agreed to work on specific sections of the report, which will be reviewed over the next week before Thursday's meeting. The council agreed to submit the interim report next Friday, April 27.

**Next meeting Thursday, April 26, 2012
1:00 – 4:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)**

**Michigan Council for Educator Effectiveness
Thursday, April 26, 2012 • 1:00 – 4:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltrown, Dave Vensel

AGENDA

1:00 – 1:30 Music teacher evaluations: Findings and recommendations

Presentation by Rick Catherman, Michigan Music Teacher Evaluation Committee chairperson, Chelsea High School director of bands, and National Board certified teacher.

To review in advance (In Dropbox): Memo from Michigan music associations (in Agendas and materials labeled Teacher Evaluation Joint Statement).

Notes: Mr. Catherman shared with council members the Michigan Music Association's recommendations for evaluating music teachers and other teachers of non-tested grades and subjects. One important consideration that Mr. Catherman noted was that music teachers should be evaluated on their ability to coordinate student musicians to play together. Unlike in other many subjects, the outcome measured has to be one of the entire group.

1:30 – 4:00 Reading, editing, and finalizing the interim progress report

Discussion led by Deborah Loewenberg Ball

Notes: The council read and reviewed the interim progress report draft. Council members made some changes, and council members were given specific writing assignments. The council plans to vote on a final draft via email tomorrow.

**Next meeting Thursday, May 10, 2012
9:00 a.m. – 12:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)**

**Michigan Council for Educator Effectiveness
Thursday, May 10, 2012 • 9:00 a.m. – 12:00 p.m.
University of Michigan**

Fleming Administration Building • 503 Thompson Street, Room 5075 • Ann Arbor

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates

9:10 – 9:20 Feedback from the Interim Progress Report submission

Notes: Council members share that feedback from the Interim Progress Report was overwhelmingly positive.

9:20 – 9:40 Review conversations with Charlotte Danielson and Rick Catherman

9:40 – 10:00 Next steps

Given the timeline we proposed in the interim progress report, how should we organize our time and thinking over the next few weeks and months? What, specifically, do we need to learn more about? What organizations or consultants do we need to hear from in council meetings?

Discussion led by Deborah Loewenberg Ball

Notes: Many next steps will depend on the pilot study. Other next steps include subgroup meetings to address non-tested grades and subjects, student growth, and waiver policies and processes.

10:00 – 10:30 Plans and considerations for the 2012-2013 pilot

Discussion led by Deborah Loewenberg Ball

Notes: There is much to plan in preparation for the pilot study. One of the biggest issues is resolving how data will be organized and who will host the data systems. The council needs to consider the pilot design, the application process, training on tools, and funding. This is the work that ISR does, but they would need additional staffing and technical systems.

10:30 – 11:30 Work in small groups

11:30 – 12:00 Reimbursement paperwork

Cori will be providing instructions and materials for council members' travel reimbursement.

**Next meeting Thursday, May 24, 2012
12:30 – 3:30 p.m.
Grand Rapids (location TBA)**

**Michigan Council for Educator Effectiveness
Thursday, May 24, 2012 • 12:30 p.m. – 3:30 p.m.
Kent ISD Building
Cedar Room • 2930 Knapp Street NE • Grand Rapids**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltrown, Dave Vensel

AGENDA

12:30 – 12:45 Opening to meeting and welcome remarks

12:45 – 1:15 Debrief Thoughtful Classroom visit in Durand, MI

Discussion led by Jennifer Hammond, principal of Grand Blanc High School in Grand Blanc, MI and **David Vensel**, principal of Jefferson High School in Monroe, MI

Notes: The Thoughtful Classroom tool can be used on an iPad or on paper. Administrators or other observers look for indicators that are labeled evident, not evident, or missed opportunity. Jenny and Dave reported that teachers and administrators seemed to like the tool and found it an effective way to provide and receive feedback.

1:15 – 2:00 Selecting observation tools for the pilot

Review observation protocols and frameworks and determine the selection process.

Discussion led by Jennifer Hammond and David Vensel

Notes: The council reviewed several tools. Suzanne offered to provide a chart or crosswalk at the next meeting that would help the council to see across tools and compare some of the characteristics of each.

2:00 – 2:30 Plans and considerations for the 2012–2013 pilot

What are our next steps for the pilot? Where do we stand with the pilot applications?

Discussion led by Deborah Loewenberg Ball, chair, Michigan Council for Educator Effectiveness; William H. Payne Collegiate Professor of Education, Arthur F. Thurnau Professor, and dean, University of Michigan School of Education

Notes: The council's next step to advance the pilot study is to send out a request for applications to all school districts in the state. These applications should be ready for the council's review at the next council meeting.

2:30 – 3:30 Public comment session

Next meeting Wednesday, June 6, 2012

9:00 – 12:00 p.m.

Ann Arbor, Fleming Administration Building, Room 4006 (503 Thompson Street)

**Michigan Council for Educator Effectiveness
Wednesday, June 6, 2012 • 9:00 a.m. – 12:00 p.m.
University of Michigan
Fleming Administration Building • 503 Thompson Street, Room 4006 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Nick Sheltrown, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates

9:10 – 9:40 Observation tool considerations

Suzanne will lead the council in considering the questions that she raised at the May 24 council meeting.

Discussion led by Suzanne Wilson

To review in advance (IN DROPBOX under Agendas and materials – June 6):

- Document titled "mcee.Observation Instrument chart.may 2012"

Notes: Suzanne reviewed the chart that she created for the MCEE that addresses the major dimensions of each of the six observation tools that the council is considering. She also examined the Michigan districts with which the tools have contracts, the training involved with each tool, and the specificity of each of the rubrics.

9:40 – 10:20 Pilot application review

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Agendas and materials – June 6):

- Document titled "Request for applications"

Notes: The council reviewed and approved the request for applications to take part in the pilot study.

10:20 – 11:00 Budget for the 2012-2013 pilot

Given the Michigan legislature's recent budgetary decisions, what is possible for a 2012–2013 pilot? What do these decisions mean for a potential contract with ISR? Given these budget constraints, what are the council's next steps?

Discussion led by Deborah Loewenberg Ball and Brian Rowan

11:00 – 12:00 Work in small groups

Next meeting Thursday, June 21, 2012

1:30 – 4:30 p.m.

Fleming Administration Building, Room 5075

**Michigan Council for Educator Effectiveness
Thursday, June 21, 2012 • 1:30 p.m. – 4:30 p.m.
University of Michigan
Fleming Administration Building • 503 Thompson Street, Room 5075 • Ann Arbor**

Council members present: Deborah Ball, Joseph Martineau, Nick Sheltroun, Dave Vensel

AGENDA

1:30 – 1:40 Opening to meeting and updates

1:40 – 2:20 Pilot timeline

Given what we understand about the budget and testing schedule, what is a feasible timeline for the pilot?

Discussion led by Deborah Loewenberg Ball

Notes: Pilot tool training will happen in August and September. ISR will conduct study in pilot districts over the course of the school year and be able to bring pilot data late in the summer of 2013.

2:20 – 3:00 Pilot design questions and discussion

What additional questions about the pilot need to be addressed to inform the design of the research study?

Discussion led by Brian Rowan

To review in advance (IN DROPBOX under Agendas and materials – June 21):

- **Document titled “mcee.pilotquestions.june2012db”**

3:00 – 3:30 Pilot application update

Cori will report on submitted applications, review districts’ questions, and discuss next steps, including district recruitment.

Discussion led by Cori Mehan

To review in advance (IN DROPBOX under Agendas and materials – June 21):

- **Spreadsheet titled “Pilot application chart”**

Notes: The council received more than 80 applications from districts across the state. Districts are asking many questions because we have not yet provided them with guidelines about the nature of the study (e.g., what they will be asked to do, when they will be trained, etc.)

3:30 – 3:50 Website updates

Discussion led by Cori Mehan

Notes: The MCEE website is up and running, but there is not yet a way for stakeholders to submit their questions and comments. The web design firm hopes to have this ready for us in July.

3:50 – 4:30 Before the July 18 meeting

What needs to be done before the July 18 meeting? How should small groups organize their time over the next four weeks?

Discussion led by Deborah Loewenberg Ball

Next meeting Wednesday, July 18, 2012 (first day of Ann Arbor Art Fairs)
8:00 a.m. – 12:00 p.m.
Fleming Administration Building

**Michigan Council for Educator Effectiveness
Thursday, July 5, 2012 • 8:00 a.m. – 12:00 p.m.
University of Michigan**

Fleming Administration Building • 503 Thompson Street, Room 4006 • Ann Arbor

Council members present: Deborah Ball, Joseph Martineau, Mark Reckase, Nick Sheltrown (via Skype), Dave Vensel

AGENDA

8:00 – 8:10 Opening to meeting and updates

8:10 – 9:00 Review observation tool training and pricing proposals

How do these proposals fit into our budget? What other questions do we have for observation tool vendors? What do districts need to know about their assigned observation tools when they are asked to take part in the pilot study?

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Agendas and materials – July 5):

- **Spreadsheet titled “Observation tool worksheets”**

Notes: The council does not yet have these proposals so this item will move to next week’s meeting.

9:00 – 9:20 Avoiding bias in the selection process

How can we manage the district selection process given the relationships that council members have with some applying school districts?

Discussion led by Suzanne Wilson

Notes: Brian is working to manage the district selection process. However, Brian anticipates that districts might not be as willing to take part in this study and that he might have to recruit, more than select, districts to gather a representative sample.

9:20 – 11:20 Application review and district selection

In order to provide pilot districts and observation tool providers enough time to prepare for the upcoming school year, the MCEE must select districts to participate in the 2012–2013 pilot and determine the observation tool that each will use.

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Agendas and materials – July 5):

- **Spreadsheet titled “Pilot application chart”**

11:20 – 12:00 Communication strategy

What is our communication strategy with the many players that are involved in this pilot (school districts, teachers, observation tool vendors, technology providers, members of the MCEE, ISR, legislators, independent observers)? Now that districts have been selected, what are the lines of communication that we must develop?

Discussion led by Deborah Loewenberg Ball

Notes: There are two types of questions that seem to come up: school-based questions about implementation and technical issues related to design. MCEE members should respond to inquiries they receive. Deborah will respond to media. Cori will respond to school districts most of the time. There needs to be more resources on the website. Council members should notify Cori with suggestions and requests for additions to the website.

Next meeting **Wednesday, July 18, 2012**
8:00 a.m. – 12:00 p.m.
Fleming Administration Building, room 4006

**Michigan Council for Educator Effectiveness
Wednesday, July 18, 2012 • 8:00 a.m. – 12:00 p.m.
University of Michigan
Fleming Administration Building • 503 Thompson Street, Room 4006 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

8:00 – 8:10 Opening to meeting and updates

8:10 – 8:40 Approval of pilot design

ISR recommends that MCEE endorse the current sampling design. This includes the approval of pilot districts and alternate districts.

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Agendas and materials – July 18):

- Document titled “MEMO on Sample for MCEE Pilot – 7_14_12”

Notes: The MCEE approves the current sampling plan, the current pilot districts, and the alternate districts.

8:40 – 9:10 Approval of pilot district recruitment and communications plan

ISR anticipates negotiation and communication challenges with school districts as we begin the recruitment and implementation phases of the pilot study. What steps can we take to ensure this process is smooth?

Discussion led by Deborah Loewenberg Ball

Notes: The council approves hiring John Graves, retired superintendent, as a consultant to work with districts during the recruitment and implementation phases of the pilot study.

9:10 – 9:50 Approval of observation tool training packages

Given the proposals from observation tool vendors and the budget we have allocated for administrator training, what training packages should we purchase for pilot districts?

Discussion led by Jenny Hammond and Dave Vensel

To review in advance (IN DROPBOX under Agendas and materials – July 18):

- Document titled “Observation tool training and price proposals”

Notes: Training for all tools should be the same; each vendor will be given four days to train administrators on the observation tools. The council has many questions for tool vendors including: does basic training include inter-rater reliability training? How is data recorded and scored? How many observations do you recommend?

9:50 – 10:50 Approval of the testing regime

Which tests will be included in the pilot? What relationships do we need to develop with these testing vendors?

Discussion led by Joseph Martineau

Notes: Joseph proposed the testing regime.

10:50 – 11:30 Data rostering systems

Discussion led by Stephanie Chardoul

Notes: Rostering systems link students to schools and to teachers. In most places where VAMs are run, roster verification happens by both teachers and administrators.

11:30 – 12:00 Upcoming meetings and next steps

The council needs to approve a schedule for fall meeting dates and discuss options for an August meeting in Grand Rapids.

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Agendas and materials – July 18):

- Document titled “Proposed_meeting_schedule_fall_2012”

Notes: The council approves the fall meeting schedule.

Next meeting Wednesday, August 15, 2012
2:00 – 5:00 p.m.
Kent Intermediate School District
2930 Knapp Street NE • Room Rogue A • Grand Rapids

**Michigan Council for Educator Effectiveness
Wednesday, August 15, 2012 • 2:00 p.m. – 5:00 p.m.
Kent Intermediate School District
2930 Knapp Street NE • Room Rogue A • Grand Rapids**

Council members present: Deborah Ball, Jenny Hammond, Mark Reckase, Nick Sheltrown, Dave Vensel (via phone)

AGENDA

2:00 – 2:10 Opening to meeting and updates

2:10 – 2:20 Financial and contractual updates

What is the status of money appropriated by the state for the pilot? What is the status of the contract between the state and ISR?

Discussion led by Deborah Loewenberg Ball

Notes: Getting money for the pilot has not been easy. The interlocal contract has not been signed by the attorney general and trainings begin next week. There are concerns about reimbursing observation tool vendors, testing costs, etc.

2:20 – 2:40 Pilot recruitment and training updates

Brian will update the council on participating pilot districts and their upcoming observation tool trainings.

Discussion led by Brian Rowan

To review in advance (IN DROPBOX under Agendas and materials – August 15):

- **Document titled “Pilot_districts_and_training_chart”**

Notes: Sampling is still fluid at this point. The biggest districts have been the hardest to recruit because many already have tools and systems that are working for them. One possibility that the MCEE discussed is allowing any of the four piloted tools to be used next year so that piloting districts will not have to switch tools.

2:40 – 3:10 Questions from pilot districts and design decisions

Brian and John will present questions from participating pilot districts for the council to discuss.

Discussion led by Brian Rowan and John Graves

To review in advance (IN DROPBOX under Agendas and materials – August 15):

- **Document titled “Questions from pilot districts”**

Notes: Many districts have questions about the costs and logistics of testing. Others have asked what happens if a pilot district's assigned tool is not chosen by the state. Some districts want to know how they will communicate with members of the MCEE. Finally, some questions have come up about administrator evaluations: will these tools be piloted and if so, which ones?

3:10 – 3:50 Contracts with pilot districts

The council will review the expectations of participating pilot districts and the agreement to participate in the pilot study.

Discussion led by Deborah Loewenberg Ball and Brian Rowan

To review in advance (IN DROPBOX under Agendas and materials – August 15):

- Documents titled “Agreement to Participate in the MCEE Pilot Program of Research” and “BR_MCEE_Responsibilities_8_10”

Notes: The most questioned expectation of participating pilot districts is the number of observations that administrators must conduct for each teacher. The council decided that this should be a minimum of three observations by the end of March.

- 3:50 – 4:30 Communication with applicant districts, pilot districts, and other stakeholders**
What is our communication strategy to: (1) notify districts that have not been selected; (2) communicate with districts that have been selected; and (3) release the names of selected districts to stakeholders across the state?

Discussion led by Deborah Loewenberg Ball

Notes: The council decided to send a letter to all districts that have not been accepted into the pilot, and then send a different letter to all districts that have been selected to participate. The council can then prepare talking points on how we recruited the districts to share with media and other stakeholders.

- 4:30 – 5:00 Next Steps**
Training for districts starts as early as next week. What are the steps that council members need to take as the pilot and pilot study begin? How should we approach the administrator evaluation tools? What are the next steps for value-added modeling?

Discussion led by Deborah Loewenberg Ball

Notes: Stakeholders are nervous about the timeline, so council members plan to reach out to teachers, the AFT, the MEA, MASSP, MASA, and MEMSPA in order to pass on news about the pilot and the council’s work. The council also begins to consider a long-term plan.

- Next meeting Wednesday, September 12, 2012**
8:00 a.m. – 10:00 a.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Wednesday, September 12, 2012 • 8:00 – 10:00 a.m.
University of Michigan School of Education
Tribute Room • 610 East University Avenue, Room 1322 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Dave Vensel

AGENDA

8:00 – 8:10 Opening to meeting and updates

8:10 – 8:35 MCEE to-do list: Edits and discussion

Deborah will share a to-do list that outlines the council's future work. What items are missing? Which items can council members work on before the next council meeting? What other changes need to be made to this document?

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Agendas and materials – September 12):

- Document titled "MCEE_TO_DO_LIST"

Notes: Council members agree that the format of the to-do list is appropriate and decide who will take on what work before the next meeting.

8:35 – 8:50 Pilot updates

Brian will provide the council with an update. What is the status of money appropriated by the state for the pilot? What is the status of training?

Discussion led by Brian Rowan

Notes: The pilot sample has been chosen and training is underway. Districts have questions about deviating from the observation tools that they were assigned. When this evaluation system is implemented statewide, there is a huge educational and technical assistance component that must be addressed. Districts need guidance on the implementation of these systems.

8:50 – 9:10 Non-tested subject teacher evaluation

In our last meeting, the council briefly discussed using subject area organizations' expertise to learn more about non-tested subject evaluation. What do we want to ask of these groups? On which subjects does the council want to focus? How would districts pilot these evaluations?

Discussion led by Deborah Loewenberg Ball

Notes: Pilot districts have so much to handle with current pilot requirements, so piloting this might be too much. Council members will look into what other states are doing with non-tested subject evaluation in order to move forward in this area.

9:10 – 9:30 Special education teacher evaluation

We have received inquiries from stakeholders regarding this specific type of educator evaluation. Does the council want to make recommendations for evaluating these teachers? How could this be incorporated into the pilot? Can special educators be evaluated using the observation frameworks tested in the pilot? How should the state measure student growth for special education students?

Discussion led by Deborah Loewenberg Ball

Notes: Council members have received numerous requests for direction and policy regarding special education evaluation. Observation tool vendors say that special education teachers can be evaluated using these tools. One example of a way to provide special education teachers with more targeted feedback might be to recommend that they be observed at least once by someone with expertise in special education.

9:30 – 10:00 Small group work

The student growth and value added subgroups will plan upcoming research and meetings.

Discussion led by Deborah Loewenberg Ball

Next meeting Thursday, October 4, 2012

1:00 – 3:00 p.m.

Detroit, MI (exact location to be determined soon)

**Michigan Council for Educator Effectiveness
Wednesday, October 3, 2012 • 3:30 – 5:30 p.m.
Garden City High School
O’Leary Auditorium • 6500 Middlebelt Road • Garden City**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

3:30 – 3:40 Opening to meeting and welcome remarks

3:40 – 4:00 Pilot update

Discussion led by Brian Rowan, Burke A. Hinsdale Collegiate Professor, University of Michigan School of Education and research professor, Institute for Social Research

Notes: All observation tool training is complete. Each tool vendor was allowed four days of training. All independent observers hired by ISR have also been trained on tools. The districts’ testing windows are opening in late October and will no longer include 12th grade testing. There are some concerns about the testing regime and the number of observations that pilot district administrators must conduct.

4:00 – 4:20 Communicating and working with stakeholders: Updates and next steps

Discussion led by Deborah Loewenberg Ball, chair, Michigan Council for Educator Effectiveness; William H. Payne Collegiate Professor of Education, Arthur F. Thurnau Professor, and dean, University of Michigan School of Education

Notes: Cori shared recent updates to the council’s website, including a function for the public to write in with questions, comments, or concerns. The council discussed other ways to engage stakeholders, including meeting with MASA, the MEA, and AFT.

4:20 – 4:40 Student growth updates

After consulting with student growth measurement vendors, this council subgroup will share the highlights of and differences between the presentations and recommendations of American Institutes for Research (AIR), Pearson, SAS, and Wisconsin Value Added Research Center.

Discussion led by Joseph Martineau, councilmember and executive director of the Bureau of Assessment and Accountability at the Michigan Department of Education, **Mark Reckase**, councilmember and professor in the measurement and quantitative methods program at the College of Education at Michigan State University, and **Nick Sheltrown**, councilmember and director of measurement, research, and accountability at National Heritage Academies in Grand Rapids

Notes: All vendors have different philosophies: some have a specific model in mind that they have found effective in other states; others are ready to work with Michigan to develop a tool. All vendors are going to be given the opportunity to work with state data as a proof of capability.

4:40 – 5:00 Administrator evaluation updates

Discussion led by Jennifer Hammond, councilmember and principal, Grand Blanc High School and **Dave Vensel**, councilmember and principal, Jefferson High School in Monroe, Michigan

Notes: Jenny and Dave discussed the current state of administrator evaluation deliberations and recommendations. They shared with the group a list of roles that would fall into the “administrator” category for teacher evaluation.

5:00 – 5:30 Public comment session

Notes: Educators shared many concerns including those regarding testing, funding the implementation of evaluations, tool training, the number of formal evaluations required, and the waiver process.

Next meeting Tuesday, October 16, 2012
12:00 – 2:00 p.m.
Lansing, MI (exact location to be determined soon)

**Michigan Council for Educator Effectiveness
Tuesday, October 16, 2012 • 12:00 – 2:00 p.m.
The Romney Building
Large Conference Room • 111 South Capitol Avenue, Floor 10 • Lansing**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

12:00 – 12:10 Opening to meeting and updates

12:10 – 12:30 Pilot update

Discussion led by Deborah Loewenberg Ball, Brian Rowan, and Stephanie Chardoul

Notes: Some pilot principals are voicing concerns about the feasibility of these evaluation tools. Schools are (1) scrambling to observe all teachers and (2) struggling to measure student growth. Principals are calling for more guidance and clarity.

12:30 – 1:00 The Education Trust-Midwest

What do the legal and regulatory frameworks for teacher evaluation look like in other states? Which parts of teacher evaluation work belong in legislation and which are better left to regulation? How does Michigan's policy landscape impact the conversation about legislation and regulation? What questions/considerations can we be thinking about now to set the MCEE's work up for success statewide?

Presentation led by **Amber Arellano**, executive director, **Drew Jacobs**, data and policy analyst, and **Sarah Lenhoff**, assistant director of policy and research, The Education Trust-Midwest.

1:00 – 1:30 Assessing and approving local evaluation tools: Planning and next steps

What are the key principles the council should use to guide recommendations around this waiver process? How do districts' growing concerns about their own evaluation tools affect our work? What are our next steps given the timeline that we have in place?

Discussion led by Deborah Loewenberg Ball

Notes: There is currently no evaluation system from which districts can opt out. Districts that have already applied for a waiver will be required to do so again after the council releases its recommendations. Issues like these have caused the council to consider pushing the entire timeline back one year, giving stakeholders time to build capacity and implement new policies.

1:30 – 1:45 Administrator evaluation updates

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

Notes: The advisory council reviewed several administrator evaluation tools, and ultimately recommended the MASA tool to the council.

1:45 – 2:00 Student growth updates

Discussion led by Mark Reckase, Nick Sheltroun, and Joseph Martineau

Notes: The council will discuss this in the next council meeting.

Next meeting Wednesday, November 7, 2012

3:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Wednesday, November 7, 2012 • 3:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

3:00 – 3:10 Opening to meeting and updates

3:10 – 3:20 Follow-up discussion from Monday's waiver conference call

Discussion led by Deborah Loewenberg Ball

Notes: It seems impossible to approve waivers before the council has made recommendations about the evaluation system. Once the council releases its final report, a group will be charged with processing and approving district waiver requests according to the council's waiver process recommendations.

3:20 – 3:40 Pilot update

Brian and Stephanie will provide a pilot update that includes feedback from the pilot district event in Lansing on October 16.

Discussion led by Brian Rowan and Stephanie Chardoul

3:40 – 4:10 Administrator evaluation updates

Jenny, Dave, and Suzanne will brief the MCEE on the Advisory Council's recommendations for administrator evaluation and on recent sub-group research and findings.

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

To review in advance (IN DROPBOX under Meeting agendas and materials – November 7):

- **Document titled: MCEE Administrator Evaluation Tool Recommendation**

Notes: Although almost all of the evaluation tools that council members are looking at cite evidence of their effectiveness, there exists no empirical evidence that any of the tools are effective. This subgroup recommends a structured review of the tools in order to gather feedback from educators about the tools.

4:10 – 4:40 Student growth updates

Mark, Nick, and Joseph will discuss progress on their work with student growth vendors. This includes both a review of upcoming work that vendors will conduct with state data this school year, as well as an update on estimates for state-wide student growth work in the 2013-2014 school year.

Discussion led by Mark Reckase, Nick Sheltrown, and Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – November 7):

- **Deliverables for the Michigan Growth Estimates**

Notes: VAM vendors provided council members with information about their work in other states. Vendors varied a lot in their approach: some seemed to have an idea of exactly the model and policies that they would employ in Michigan, while others were more interested in an interactive approach, adapting practices based on stakeholder feedback.

4:40 – 5:00 Student Learning Outcomes (SLOs): One way to measure student growth in non-tested subjects

Discussion led by Jennifer Hammond

To review in advance (IN DROPBOX under Meeting agendas and materials – November 7):

- Student Learning Objectives Handbook Version 2.0_Indiana
- Student Learning Objectives Guide for Educators - Writing SLOs_RI
- Rhode Island SLO FAQ

Notes: Did not have time for this; moved to next meeting.

Next meeting Monday, November 26, 2012
4:00 – 6:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Monday, November 26, 2012 • 4:00 – 8:00 p.m.
University of Michigan School of Education
Brownlee Room • 610 East University Avenue, Room 2327 • Ann Arbor**

Council members present: Deborah Ball, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

4:00 – 4:10 Opening to meeting and updates

4:10 – 4:20 Pilot update

Discussion led by Brian Rowan and Stephanie Chardoul

Notes: The fall testing regime is now complete and the paired observations in pilot districts are going on now. The next step is to analyze the data from these observations. Also, ISR has created the student-teacher rostering system in which each teacher will be using a secure web-portal to record his or her class rosters.

4:20 – 4:40 Recommendation timeline updates

According to the MCEE timeline published in the interim progress report, several recommendation deadlines are approaching or have already passed. What are reasonable goals the council can set and immediate actions that we can take looking toward the end of this year and the beginning of 2013?

Discussion led by Deborah Loewenberg Ball

Notes: The council's next task is to make student growth recommendations. Joseph will try to write clear language to distinguish between student growth models and growth and assessment tools.

4:40 – 5:00 Educator evaluation tool training: Statewide implementation and the state budget

Discussion led by Jennifer Hammond

Notes: Jenny suggested that the council should collect from tool vendors their plans for statewide implementation of their products, including pricing and training at scale.

5:00 – 5:30 Administrator evaluation: Next steps

Jenny, Dave, and Suzanne will outline their recommendations for the next steps in the administrator evaluation process, including plans for a structured review by administrators in several districts across the state.

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

Notes: Each focus group will look at two administrator evaluation tools, compare them, and provide feedback based on the questions that the council creates. ISR has the capacity and the budget to organize and conduct these groups and will work with Suzanne to post candidate invitations and to provide candidates with a list of questions.

5:30 – 6:15 Classifying teachers, combining scores, and calculating error

Discussion led by Brian Rowan

Notes: Teacher observations and VAMs are imprecise. To improve precision on the practice portion of teacher evaluations, teachers should be observed three to four different times by different people. VAMs' precision increases with more numbers.

6:15 – 7:00 Student growth and value-added modeling discussion

How should the council recommend that value-added data be used for educator evaluation next year?

Discussion led by Deborah Loewenberg Ball

Notes: Moved to next meeting due to time.

7:00 – 7:30 Student Learning Outcomes (SLOs): One way to measure student growth in non-tested subjects

Discussion led by Jennifer Hammond

To review in advance (IN DROPBOX under Meeting agendas and materials – November 26):

- Student Learning Objectives Handbook Version 2.0_Indiana
- Student Learning Objectives Guide for Educators - Writing SLOs_RI

Notes: Moved to next meeting due to time.

7:30 – 8:00 Waiver process

What have other states done to build guidelines and manage district exemptions? What are the next steps we can take toward completing these waiver process recommendations?

Discussion led by Suzanne Wilson and Cori Mehan

Notes: Suzanne presented a waiver memo that she shared with the council. Other states' waiver policies seem to fall on a spectrum: Rhode Island has a very complex system of standards and rubrics; Indiana has a checklist and allows the state to make compliance decisions; Massachusetts simply requires districts to answer questions about their own evaluation systems.

Next meeting Monday, December 11, 2012
11:00 a.m. – 3:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Tuesday, December 11, 2012 • 11:00 a.m. – 3:00 p.m.
University of Michigan School of Education
Room 2334 • 610 East University Avenue • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

11:00 – 11:10 Opening to meeting and updates

11:10 – 11:20 Pilot update

Discussion led by Stephanie Chardoul

Notes: The first round of paired observations is still happening in pilot schools. ISR is working with all four observation vendors to develop a calibration system for their researchers/observers to complete while schools are closed for the holiday break, including viewing and scoring of master coded video.

11:20 – 11:50 Student growth updates

Mark, Nick, and Joseph will discuss progress on their work with student growth vendors. This includes a review of upcoming work that vendors will conduct with state data this school year and an explanation of how that work will help the council make fair, transparent, and feasible student growth recommendations.

Discussion led by Mark Reckase, Nick Sheltrown, and Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – December 11):

- **Deliverables for the Michigan Growth Estimates**

Notes: This subgroup is planning to ask VAM vendors to do work for us in order to see how they might work with state data. This is an unusual opportunity because the council will have four different analyses of the same material and will be able to look across vendors in order to see their analysis, how they plan reporting, and how they manage to combine files.

11:50 – 12:50 Student growth and value-added modeling discussion

How should the council recommend that value-added data be used for educator evaluation next year?

Discussion led by Deborah Loewenberg Ball

Notes: The council is aware that all methods of evaluation—VAM, teacher observation, and total administrator discretion—are all prone to error. Though it might seem that VAM will not be able to inform a teacher of how to improve, it will provide for teachers and administrators evidence of growth.

12:50 – 2:00 Non-tested grades and subjects and Student Learning Objectives (SLOs)

Given the Michigan Department of Education's plans to create tests for all non-tested grades and subjects, how should the council approach this part of our charge? What is a fair and feasible plan for the interim period before these tests are available?

At 1:30, we will have a Skype conversation with **James H. Stronge**, Heritage Professor in the Educational Policy, Planning, and Leadership Area at the College of William and Mary, regarding his research related to assessing teachers in non-tested grades and subjects.

Discussion led by Jennifer Hammond

To review in advance (IN DROPBOX under Meeting agendas and materials – December 11):

- Student Learning Objectives Handbook Version 2.0_Indiana
- Student Learning Objectives Guide for Educators - Writing SLOs_RI
- Stronge Sample Goal Setting Guidebook

Notes: SLOs can be done at any level and in any subject area. They allow teachers to create a rigorous goal that they intrinsically value and allow for agreed upon standards between teachers and principals. Implementing SLOs requires laying down ground rules, creating a guide book that tells districts, principals, and teachers how to do this, developing a library of good SLOs, and training educators. Council members agree that SLOs would generate a lot of educator buy-in and seem like a possible method of evaluation to incorporate into recommendations.

2:00 – 2:20 Recommending other components of teacher evaluation systems

Does the council want to recommend other components of teacher evaluation systems, such as student surveys, parent feedback, and peer observations, or should these decisions be left up to local districts?

Discussion led by Deborah Loewenberg Ball

Notes: Nick suggests that it does not seem like the council has time to research and recommend any of these given our timeframe. Joseph thinks, however, that it seems important to give districts some room to add evaluation practices of their own.

2:20 – 3:00 Work in small groups

Next meeting Thursday, December 20, 2012
3:00 p.m. – 5:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Thursday, December 20, 2012 • 3:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

3:00 – 3:10 Opening to meeting and updates

3:10 – 3:30 Pilot update and observation tool vendor RFP discussion

Discussion led by Brian Rowan and Stephanie Chardoul

To review in advance (IN DROPBOX under Meeting agendas and materials – December 20):

- Questions for observation tool vendors

Notes: Stephanie and Brian described the complexity of the roster verification process. They suggested that the state develop and maintain a uniform data system that incorporates many systems.

3:30 – 3:50 Next steps for the administrator evaluation structured review

Discussion led by Jenny Hammond, Dave Vensel, and Suzanne Wilson

Notes: The council reviewed Suzanne's structured review document. Deborah suggested the inclusion of the purpose and use of the data we collect. Stephanie and Brian requested a more developed memo so that they can understand (1) the council's exact questions and (2) the way that the council wants results reported.

3:50 – 4:40 Measuring effectiveness: Observation tools and student growth measures

As a continuation of our conversation in last week's meeting, how are council members thinking about the measures of effectiveness that observation protocols, value-added models, SLOs, and other tools provide? Given the uncertainty with which we can categorize teachers using many of the methods we have learned about, how should the council approach final recommendations?

Discussion led by Deborah Loewenberg Ball

Notes: The council agreed to lay out the recommendations that they could implement now, then what could be put into place in 2014-15, and finally what could happen in 2015-16. This would be like providing the state a road map for implementation.

4:40 – 5:00 First steps for 2013

What work should council members and subgroups focus on between now and our next meeting? What are the council's next steps when we return from the holiday break? What is our plan and timeline to review and post our update?

Discussion led by Deborah Loewenberg Ball

Next meeting Friday, January 11, 2013

3:30 – 5:30 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Friday, January 11, 2013 • 3:30 – 5:30 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

3:30 – 3:40 Opening to meeting and updates

3:40 – 3:50 Pilot update

Discussion led by Brian Rowan and Stephanie Chardoul

Notes: ISR has been working with schools to schedule the second round of paired observations. Also, ISR is pursuing additional contracts with observation vendors to provide an inter-rater reliability check and additional training for principals who would like to participate.

3:50 – 4:10 Student growth rubrics

Discussion led by Jennifer Hammond

To review in advance (IN DROPBOX under Meeting agendas and materials – January 11)

- wa-student-growth-rubrics

Notes: Council members reviewed the student-growth rubrics that Jenny provided. Council members were concerned that the rubrics did not define “growth” or “high growth,” and that the use/results of the districts would vary widely from district to district.

4:10 – 4:40 Student growth and assessment tools: Draft review

After reviewing Joseph's draft of the MCEE's recommendations for student growth and assessment tools, what questions and suggestions do council members have?

Discussion led by Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – January 11)

- DRAFT recommendations for student growth and assessment tools

Notes: Some questions that came from Joseph's presentation were: how will teachers be categorized? How will SLOs, VAMs, and observations all fit together? In recommendations, how will we connect for our audience growth tools and evaluation?

4:40 – 5:00 Budget discussion

Discussion led by Dennis Schornack

5:00 – 5:30 Structuring final recommendations: A continuing conversation

In our last two meetings, we discussed recommending a phase-in of evaluation tools in order to ensure feasible and effective implementation. We also discussed the structure of the final report. What additional ideas about a phase-in or about the structure of the final report do council members have? As Brian asked in our last meeting, will we be recommending a locally controlled system or a state system? What are the benefits and disadvantages of both?

Discussion led by Deborah Loewenberg Ball

Notes: The council developed a list of questions that must be answered the report.

Next meeting **Friday, February 1, 2013**
10:00 a.m. – 1:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Friday, February 1, 2013 • 10:00 a.m. – 1:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

10:00 – 10:10 Opening to meeting and updates

10:10 – 10:20 Pilot update

Discussion led by Stephanie Chardoul

10:20 – 11:00 Structuring our final recommendations

Discussion led by Deborah Loewenberg Ball

Notes: The council continued their discussion about their final recommendations. One idea the council discussed was creating an implementation year to provide the state enough time to build capacity and employ the resources needed to effectively implement recommendations.

11:00 – 12:30 Reviewing subgroup work and recommendations

Subgroups have been meeting over the past two weeks to address the key policy questions that the council drafted in our last meeting. What recommendations came from these meetings? What questions does the whole group still need to address?

Discussion led by Jennifer Hammond

To review in advance (IN DROPBOX under Meeting agendas and materials – February 1)

- Jan30_value-added_subgroup_memo
- Jan30_reporting and ratings_subgroup_memo
- Jan30_teacher and administrator evaluations_subgroup_memo
- Jan30_logistics and implementation_subgroup_memo

Notes: Subgroups summarized their individual discussions and posed recommendation questions to the whole group for continued discussion and consideration.

12:30 – 1:00 Developing a communications plan

What steps should the council take to build support for our final recommendations? What are the key stakeholder groups to be engaged in this work (e.g., teachers, administrators, policymakers, parents, the business community)? How will the council communicate with these stakeholders before and after the release of our final report? Are there groups that we want to enlist to help us build consensus among these stakeholders?

Discussion led by Deborah Loewenberg Ball

Notes: Administrators, teachers, parents, legislators, and others involved in the Michigan education community all need to be engaged in this work. Council members discussed the idea of using conferences and other speaking engagements to educate stakeholders about the council's recommendations and vision for evaluation in Michigan.

Next meeting **Friday, February 8, 2013 – CANCELED BECAUSE OF WEATHER**
9:00 a.m. – 12:00 p.m.
The Skillman Foundation (100 Talon Centre Dr. Suite 100, Detroit)

**Michigan Council for Educator Effectiveness
Monday, February 18, 2013 • 2:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown (via phone), Dave Vensel

AGENDA

2:00 – 2:10 Opening to meeting and updates

2:10 – 2:20 Pilot update

Notes: ISR continues to conduct 2nd round teacher observations. ISR is also continuing with calibration activities for school researchers across all four observation protocols. This involves viewing videos that have been "master coded" by the vendors, coding, and then comparing and discussing results. They have also finalized the testing plans and dates for spring assessments.

2:20 – 3:00 Reviewing subgroup work and recommendations

Discussion led by Jennifer Hammond and Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – February 18)

- Jan30_memo_implementation
- Joseph's weights and measures memo (will be in Dropbox on Friday)

Notes: Joseph recommended that the council have a statewide VAM model so that we can have help combining scores. However, the council needs to work to build the capacity of likely players so that we will have wiser decision makers.

3:00 – 3:30 Speaking with American Institutes for Research

Over the past few months, the council has worked to gather information about value-added modeling and other measures of student growth. As we prepare to draft final recommendations, council members still have important questions about how value-added modeling will account for practitioners' specific instructional situations and how this information will help them to improve their practice.

Discussion led by Nick Sheltrown

Notes: AIR answered council members' questions about student teachers, missing data, co-teachers, and long-term absences. AIR was also able to provide the council with information about educating teachers, administrators, and parents about VAMs and the information they provide.

3:30 – 4:30 Drafting our final recommendations

Given our previous discussions and council members' work that we reviewed in our last meeting, what are our next steps toward drafting final recommendations?

Discussion led by Deborah Loewenberg Ball

Notes: Deborah proposes that we use the “require, permit, encourage, and prohibit” framework. Many districts are mobilizing in anxiety so the council needs to recommend a year delay in the implementation process. This will provide time issuing RFPs, developing contracts, and finalizing the new testing regime. The motion to delay is now in Dropbox.

4:30 – 5:00 Developing a communications plan

- What steps should the council take to build support for our final recommendations?
- What are the key stakeholder groups to be engaged in this work (e.g., teachers, administrators, policymakers, parents, the business community)?
- How will the council communicate with these stakeholders before and after the release of our final report?
- Are there groups that we want to enlist to help us build consensus among these stakeholders?

Discussion led by Deborah Loewenberg Ball

Notes: Move this discussion to the next meeting.

Next meeting Friday, March 15, 2013
1:00 – 4:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Friday, March 15, 2013 • 1:00 – 3:00 p.m.
University of Michigan School of Education
Teaching Works Conference Room • 610 East University Avenue, Room 1032 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Dave Vensel

AGENDA

1:00 – 1:10 Opening to meeting and updates

1:10 – 1:20 Pilot update

Notes: March marked the end of round 3 observations, and Stephanie believes that most districts are on the same schedule. ISR is negotiating a lot of back and forth to get usable data from the vendors. Stephanie meets with Brian on Monday and they are setting deadlines for reports and data to the MCEE.

1:20 – 1:40 Reviewing and approving student growth recommendations

Discussion led by Mark Reckase

To review in advance (IN DROPBOX under Meeting agendas and materials – March 15)

- Student growth recommendations_March15 (will be in Dropbox Friday morning)

Notes: There is some concern about the testing regime and a call for carefulness around our recommendations. Are we talking about tests that the state will have, or tests that the state already has? This subgroup also raised the issue of combining different components of student growth to reach a student growth score.

1:40 – 2:00 Approving waiver process recommendations

Discussion led by Jenny Hammond and Suzanne Wilson

Notes: Districts can opt out of either or both the teacher and administrator evaluation systems. Criteria for opting out are feasible for districts, but not easy for districts. Districts will have three years to prove reliability and validity of their local systems. More details about waivers can be included in an “infrastructure” section of the final report.

2:00 – 3:00 Crafting the final report

The council will discuss resolved and unresolved recommendations.

Discussion led by Deborah Loewenberg Ball

Next Meeting April 10, 2013

2:00 – 5:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Wednesday, April 10, 2013 • 2:00 – 5:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

2:00 – 2:10 Opening to meeting and updates

2:10 – 2:20 Pilot update

Notes: Paired observations are wrapping up and the spring testing period is going to launch soon. ISR is still working with MDE and CEPI to retrieve the requested data. ISR has finished the rostering system and is now cleaning and formatting the rosters. The administrator evaluation structured review is complete and ISR has completed their report.

2:20 – 3:00 Implementing teacher evaluation in Michigan

After the council releases its final report, what group or groups will ensure that the system is implemented? Where should this office be housed? What role will the current council play in future teacher evaluation work?

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

Notes: This subgroup argued that implementing this system is a public education campaign. Jenny, Dave, and Suzanne suggest that having an organization located within MDE, as well as having an independent body made up of practitioners, will both be very valuable to implementation.

3:00 – 5:00 Crafting and editing the final report

The council will continue to discuss resolved and unresolved recommendations and to begin to edit the final report draft.

Discussion led by Deborah Loewenberg Ball

Notes: Deborah assigned council members to work on specific sections of the final report draft.

Next meeting Monday, April 22, 2013

9:00 – 11:30 a.m.

The Romney Building, Constituent Relations Conference Room (111 S. Capitol Ave., Lansing)

**Michigan Council for Educator Effectiveness
Monday, April 22, 2013 • 9:00 – 11:30 a.m.
The Romney Building**

Constituent Relations Conference Room • 111 South Capitol Avenue, Floor 1 • Lansing

Council members present: Deborah Ball, Joseph Martineau, Mark Reckase, Nick Sheltroun, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates

9:10 – 9:40 Implementing teacher evaluation in Michigan

After the council releases its final report, what group or groups will ensure that the system is implemented? Where should this office be housed? What role will the current council play in future teacher evaluation work?

Discussion led by Dave Vensel and Suzanne Wilson

Notes: Some of the functions of the system supporting the evaluation system will be gathering information, keeping track of tools and instruments, data tracking and storage, capacity building and training, and approving waivers and compliance. Some core questions include: should all of these functions be in one office? Where should this office be located? How should resources be allocated? Should the MCEE have a role to oversee various functions?

9:40 – 10:10 Categorizing teachers

What is accomplished by rating teachers? If we do place teachers into categories, what are the categories we want to recommend? How would the council define or describe these categories?

Discussion led by Deborah Loewenberg Ball

Notes: Some teachers could be rated more accurately than others. Statistically, it is easier to identify the very strong and the very weak teachers. It is more difficult to identify teachers in the middle.

10:10 – 10:50 Speaking with the Center for Educational Performance and Information (CEPI)

The council will be speaking with **Thomas Howell**, director of CEPI, who is responsible for directing CEPI's efforts as they relate to individual student and staff data, school finance and safety data, and other data collection and feedback efforts. He is also responsible for the implementation of the state's education longitudinal data system, and coordinates overall policy, planning and administrative efforts at CEPI. **Trina Anderson**, assistant director of CEPI, will also be joining us.

Discussion led by Joseph Martineau

10:50 – 11:20 Reviewing assessment tool memos and recommendations

Discussion led by Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – April 22)

- DRAFT MCEE recommendations re growth and assessment tool
- DRAFT Recommendations for Using Student Growth and Achievement Data in Educator Evaluations

Notes: Joseph led the council through these drafts. The council discussed making the language about the complexities of value-added accessible to stakeholders without oversimplifying these measures.

11:20 – 11:30 Updates and assignments for the final report

Discussion led by Deborah Loewenberg Ball

Next meeting Wednesday, May 8, 2013

9:00 a.m. – 12:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Wednesday, May 8, 2013 • 9:00 a.m. – 12:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates

9:10 – 9:20 Pilot update

Notes: May 10 is the deadline to have all data in the system for the final paired observations. ISR is still working with the state to get student data.

9:20 – 10:00 Reviewing student growth and assessment tool recommendations

Discussion led by Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – May 8)
• final_report_draft_May_8_jm (section on student growth and assessment tools)

Notes: Recommending a student growth measurement layer is important, just as it is important that we recommend tools for assessing teacher and administrator practice. We also need to provide specific enough guidance to the legislature so that the assessments they adopt are aligned with standards. The council agreed on the importance of using strong, clear language to talk about what student growth actually is and the system's ability to use student growth.

10:00 – 10:30 Reviewing administrator evaluation recommendations

After reading ISR's report on the administrator evaluation structured review and collaborating with ISR survey director Barbara Ward, this subgroup worked to draft recommendations for the council's final report.

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

To review in advance (IN DROPBOX under Meeting agendas and materials – May 8)
• DRAFT_administrator_evaluation_memo_4_26

Notes: All administrators seemed to agree that simpler systems are better, but a clear "best" tool did not emerge. There was some concern that the varying roles of administrators were not fully represented in the tools, so some wanted to be able to adapt tools to fit administrator's roles.

10:30 – 11:00 Infrastructure and implementation recommendations and next steps

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

Notes: The council will work more on this at the next meeting.

11:00 – 11:40 Continuing our conversation: Categorizing teachers

What is accomplished by rating teachers? If we do place teachers into categories, what are the categories we want to recommend? How would the council define or describe these categories?

Discussion led by Deborah Loewenberg Ball

Notes: Deborah presents the three-category chart to council members. If our goal is improvement for everyone, what is the goal of the “highly effective” category? Eliminating it would prevent unnecessary competition among teachers and the idea that no improvement is needed.

11:40 – 12:00 Updates and assignments for the final report

Discussion led by Deborah Loewenberg Ball

Next meeting Wednesday, May 15, 2013

1:00 – 4:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Wednesday, May 15, 2013 • 1:00 – 4:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

1:00 – 1:10 Opening to meeting and updates

1:10 – 1:40 Reviewing revised student growth and assessment tool recommendations

Discussion led by Joseph Martineau

To review in advance (IN DROPBOX under Meeting agendas and materials – May 15)
• final_report_draft_May_15 (section on student growth and assessment tools)

Notes: The state will provide tests, but will not require districts to use that data for teacher evaluations. Then, the state will continue to grow and develop with the system, which is still in its infancy.

1:40 – 2:00 Reviewing revised administrator evaluation recommendations

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

To review in advance (IN DROPBOX under Meeting agendas and materials – May 15)
• DRAFT_administrator_evaluation_memo_May15

Notes: We will do this in the next meeting.

2:00 – 2:30 Infrastructure and implementation recommendations and next steps

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

Notes: Jenny, Dave, and Suzanne outlined the positions and employees needed to make up the infrastructure of the evaluation system, most notably the “evaluation czar”. The czar would be a temporary position to get this system up and running and to meet with and coordinate among stakeholders.

2:30 – 3:00 Professional certificate recommendations

Discussion led by Dave Vensel

Notes: The big criticism of the current system is that requirements include time in classes and professional development but do not include many outputs. The council must decide what it wants to take on and what, at this point, there is time to recommend.

3:00 – 3:30 Continuing our conversation: Categorizing teachers

Given our discussion in last week's meeting, how can we begin to think about categorizing teachers without using the “highly effective” category? What are alternative labels that could better describe categories?

Discussion led by Deborah Loewenberg Ball

Notes: Not having a “highly effective” category makes it difficult to have a meritorious category; however, this category makes evaluation a competitive issue. As far as policy is concerned, it is better to be able to identify the lowest group of teachers.

3:30 – 4:00 Updates and assignments for the final report

Discussion led by Deborah Loewenberg Ball

**Next meeting Thursday, May 30, 2013
1:00 – 4:00 p.m.
The Romney Building (111 S. Capitol Ave., Lansing)**

**Michigan Council for Educator Effectiveness
Thursday, May 30, 2013 • 1:00 – 4:00 p.m.
Michigan Department of Education (The Hannah Building)
Ottawa 2 Room • 608 West Allegan Street • Lansing**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

1:00 – 1:10 Opening to meeting and updates

1:10 – 2:00 Reviewing revised student growth and assessment tool recommendations

Discussion led by Joseph Martineau, Nick Sheltrown, and Dave Vensel

To review in advance (IN DROPBOX under Meeting agendas and materials – May 30)

- student growth and assessment tools section (5_28_2013)

Notes: Interim tests will only be in core subjects; districts will be able to choose to use them for VAMs and it will be optional for districts to use the VAMs for evaluation. The council needs to use expertise, efficiency, and latitude in order to make recommendations that enable flexibility as well as promote the smartest policies.

2:00 – 2:30 Reviewing revised administrator evaluation recommendations

Discussion led by Jennifer Hammond, Dave Vensel, and Suzanne Wilson

To review in advance (IN DROPBOX under Meeting agendas and materials – May 30)

- DRAFT_administrator_evaluation_memo_May30

2:30 – 3:00 Professional certificate discussion

In last week's meeting, the council discussed the potential breadth of this portion of the charge. How should we approach and focus these recommendations given our limited timeframe?

Discussion led by Dave Vensel

Notes: Our country spends a lot of money on professional development, but it is difficult to find something that is relevant and high quality. The problem with providing local districts flexibility around this issue is that the system is fraught with a lack of quality. The council decided to recommend explicitly that there is more work to be done before we can tie preparation to effectiveness, but we can say what should not be done.

3:00 – 3:40 Continuing our conversation: Categorizing teachers

Given our discussion in last week's meeting, how can we begin to think about categorizing teachers without using the "highly effective" category? What are alternative labels that could better describe categories?

Whatever we recommend about labels, how are we recommending that different information be combined to arrive at a rating?

What are some of the legal considerations, given what we have learned about the reliability of ratings?

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Meeting agendas and materials – May 30)

- graves legal considerations_may28

Notes: Mark worries about the accuracy of classifying people. What are the consequences of not being able to identify marginally effective teachers?

3:40 – 4:00 Updates and assignments for the final report

Discussion led by Deborah Loewenberg Ball

Next meeting Monday, June 10, 2013
9:00 a.m. – 12:00 p.m.
University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Monday, June 10, 2013 • 9:00 a.m. – 12:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond, Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

9:00 – 9:10 Opening to meeting and updates

9:10 – 9:50 Making final decisions: Growth and assessment tool recommendations

Discussion led by Suzanne Wilson

To review in advance (IN DROPBOX under Meeting agendas and materials – June 10)
• student growth and assessment tools section (6_10_2013)

Notes: This is not ready yet. Deborah and Suzanne will continue to improve this section and will have a draft for council members at next week's meeting.

9:50 – 10:20 Making final decisions: Administrator evaluation recommendations

Discussion led by Jennifer Hammond, Dave Vensel, Joseph Martineau, and Suzanne Wilson

To review in advance (IN DROPBOX under Meeting agendas and materials – June 10)
• DRAFT_administrator_evaluation_memo

Notes: Like teacher evaluation, the council is also proposing three categories for administrator evaluation. The council also needs to recommend that data be collected over time on the administrator evaluation tools and data because there is little evidence about the validity and reliability of these tools.

10:20 – 10:50 Categorizing teachers and combining scores

Given our decision to recommend three categories, how are we recommending that different information be combined to arrive at a rating?

Discussion led by Deborah Loewenberg Ball

Notes: Increasing the size of the ineffective category is not useful; there are actually very few teachers that fall into this group. Districts and principals need to be allowed to use tools (SLOs, VAM scores, and observation data) to inform decisions.

10:50 – 11:20 Surfacing perceived conflicts of interest

As we discussed in last week's meeting, there are several perceived conflicts of interest between council members and consultants and the council's recommendations. What are these conflicts of interest and how can we adapt our messaging to manage potential stakeholder concern?

Discussion led by Deborah Loewenberg Ball

Notes: The council will address this in the next meeting.

11:20 – 11:50 Presentation by Pace and Partners

Representatives **Roni Rucker Waters** and **Chelsea Maupin** from Pace will guide the council through a tentative communications plan that covers preparations for the final report release and the marketing campaign to follow.

Notes: Pace shared their communications plan with the council, which includes town halls, a media press conference, and other opportunities for council members to reach out to stakeholders to increase understanding about the recommendations.

11:50 – 12:00 Planning and assignments for finishing the report

Discussion led by Deborah Loewenberg Ball

Next meeting Monday, June 17

3:00 – 6:00 p.m.

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

**Michigan Council for Educator Effectiveness
Monday, June 17, 2013 • 3:00 – 6:00 p.m.
University of Michigan School of Education
Dean's Conference Room • 610 East University Avenue, Room 1211 • Ann Arbor**

Council members present: Deborah Ball, Jenny Hammond (via phone), Joseph Martineau, Mark Reckase, Nick Sheltrown, Dave Vensel

AGENDA

3:00 – 3:10 Opening to meeting and updates

3:10 – 4:00 Making final decisions: Growth and assessment tool recommendations

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Meeting agendas and materials – June 17)
• student growth and assessment tools section (6_17_2013)

Notes: The council decided to recommend that the state will have standards and a framework, and that the state will have tests that measure state standards.

4:00 – 5:00 Categorizing teachers and combining scores

Given our decision to recommend three categories, how are we recommending that different information be combined to arrive at a rating?

Discussion led by Deborah Loewenberg Ball

To review in advance (IN DROPBOX under Meeting agendas and materials – June 17)
• Possible Conceptual Frameworks for a Three Category System

Notes: The council discussed qualitative vs. quantitative combining options, but decided that a chart (qualitative calculation) will be best. The council also decided on clear definitions for the three rating categories.

5:00 – 5:40 Surfacing perceived conflicts of interest

As we discussed in last week's meeting, there are several perceived conflicts of interest between council members and consultants and the council's recommendations. What are these conflicts of interest and how can we adapt our messaging to manage potential stakeholder concern?

Discussion led by Deborah Loewenberg Ball

Notes: The council will do this in the next meeting once the draft is complete.

5:40 – 6:00 Planning and assignments for finishing the report

Discussion led by Deborah Loewenberg Ball

**Next meeting Tuesday, June 25
8:00 – 11:00 a.m.**

University of Michigan School of Education (610 E. University Avenue, Ann Arbor)

Appendix C: In-Meeting Consultations

Name	State/ Position	Date Consulted	Information Provided
Trina Anderson	Michigan Assistant Director of the Center for Performance and Information (CEPI)	April 22, 2013	Ms. Anderson shared with the council information about student rostering systems and the linking of teachers to students to assess instructional responsibility.
Mark Atkinson	Founder and CEO of Teachscape	April 19, 2012	Mr. Atkinson explained and demonstrated Teachscape's online training portal for the Danielson Framework for Teaching (2011).
Sandy Austin	State of Washington Project director at the Center for Education Leadership, University of Washington College of Education	March 7, 2012	Ms. Austin contributed to the presentation on the Five Dimensions framework, which was built in the University of Washington's College of Education.
Rick Catherman	Michigan Michigan Music Teacher Evaluation Committee chairperson, Chelsea High School director of bands, and National Board certified teacher	April 26, 2012	Mr. Catherman explained his findings regarding music teacher evaluations, and made recommendations for addressing non-tested subject evaluations.
Beth Carr	Director of District Partnerships, Learning Sciences International	March 20, 2012	Ms. Carr helped the council to learn more about the implementation, feasibility, and training of the Marzano Causal Teacher Evaluation Model.
Stephanie Chardoul	Michigan Survey director, Institute of Social Research, University of Michigan	Ongoing	Ms. Chardoul worked to organize, implement, and analyze data from the 2012–2013 Michigan Pilot of Educator Evaluation Tools and frequently reported pilot progress to the MCEE.

Name	State/ Position	Date Consulted	Information Provided
Charlotte Danielson	Founder of the Danielson Group and creator of the Danielson Framework	April 19, 2012	Ms. Danielson explained more details about training, cost, feasibility, reliability, and validity of the Danielson Framework.
Julie Everly	Michigan Assistant superintendent for elementary education, Monroe Public Schools	January 18, 2012	Ms. Everly explained and answered questions about the iPad walk-through tool now used in Monroe Public Schools.
Steve Fink	State of Washington Executive director at Center for Education Leadership, University of Washington College of Education	March 7, 2012	Mr. Fink contributed to the presentation on the Five Dimensions framework, which was built in the University of Washington's College of Education.
John Graves	Michigan Lecturer, University of Michigan School of Education and former superintendent, Jackson County Intermediate School District	Ongoing	Mr. Graves worked closely with the 2012–2013 Michigan Pilot Study of Educator Evaluation Tools and provided guidance to the council regarding the implementation of evaluation tools in districts across the state.
Edie Holcomb	State of Washington Program facilitator at Washington's Teacher and Principal Evaluation Pilot	March 7, 2012	Mr. Holcomb contributed to the presentation on the Five Dimensions framework, which was built in the University of Washington's College of Education.
Thomas Howell	Michigan Director of the Center for Performance and Information (CEPI)	April 22, 2013	Mr. Howell is responsible for the implementation of the state's education longitudinal data system. He shared with the council information about student rostering systems and the linking of teachers to students to assess instructional responsibility.

Name	State/ Position	Date Consulted	Information Provided
Max Hunsicker	Michigan Senior director of coaching and learning, National Heritage Academies	March 1, 2012	Mr. Hunsicker shared information regarding National Heritage Academies' teacher evaluation system.
Jason Kennedy	Michigan Principal, Oakridge Public Schools	February 27, 2012	Mr. Kennedy discussed the evaluation system currently used by Oakridge Public Schools.
John Lane	Michigan Researcher and graduate student, Michigan State University	Ongoing	Mr. Lane attended several council meetings and provided targeted assistance and research to council members.
Tom Livezey	Michigan Superintendent, Oakridge Public Schools	February 27, 2012	Mr. Livezey discussed the evaluation system currently used by Oakridge Public Schools.
Dan McCaffrey	PNC Chair in Policy Analysis and senior statistician at RAND Corporation	February 21, 2012	Mr. McCaffrey guided the council through an introduction of value-added modeling and answered council members' questions.
Laurie McCullough	Chief Strategy Officer, Teachstone	March 20, 2012	Ms. McCullough helped the council to learn more about the implementation, feasibility, and training of CLASS observation tool.
Ryan McLeod	Michigan Assistant superintendent for secondary education, Monroe Public Schools	January 18, 2012	Mr. McLeod explained and answered questions about the iPad walk-through tool now used in Monroe Public Schools
Michaela Miller	State of Washington Program manager, Washington's Teacher and Principal Evaluation Pilot	April 2, 2012	Ms. Miller shared information regarding the training, cost, and feasibility of an evaluation tool pilot, like the one she is working with in Washington.

Name	State/ Position	Date Consulted	Information Provided
Brian Rowan	Michigan Burke A. Hinsdale Collegiate Professor, University of Michigan School of Education	Ongoing	Dr. Rowan attended many council meetings as an ongoing consultant. He has provided guidance around student growth modeling, calculating validity and reliability, assessment, and understanding large-scale implementation of evaluation tools in schools and LEAs.
Lorrie Shepard	Colorado Dean & Distinguished Professor, School of Education, University of Colorado at Boulder	March 16, 2012	Dean Shepard met with council members via Skype to explain the educator evaluation reform process in Colorado. She discussed Colorado's timeline, resources, process, and lessons learned.
James Stronge	Heritage Professor in the Educational Policy, Planning, and Leadership Area at the College of William and Mary.	December 11, 2013	Dr. Stronge shared his research and expertise related to assessing teachers in non-tested grades and subjects, including the use of student learning objectives (SLOs).
Ginny Vitello	Research and evaluation director, Teachstone	March 20, 2012	Ms. Vitello helped the council to learn more about the implementation, feasibility, and training of CLASS observation tool.
Barbara Ward	Michigan Survey director, Institute of Social Research, University of Michigan	May 8, 2013	Ms. Ward worked to organize and gather and analyze data from the structured review of four administrator evaluation tools. Ms. Ward presented results to the council.

Name	State/ Position	Date Consulted	Information Provided
Suzanne Wilson	Michigan University Distinguished Professor, chair of the department of Teacher Education, and director of the College of Education's Center for the Scholarship of Teaching, Michigan State University	Ongoing	Dr. Wilson attended many council meetings as an ongoing consultant. She provided invaluable information regarding observation tools, other states' experiences, and the large-scale implementation of evaluation systems in schools and LEAs. Dr. Wilson has also written memorandums that helped to organize and articulate the council's ideas and findings.

Appendix D: Out-of Meeting Consultations

Name	Position	Date Consulted	Information Provided
Katy Anthes	Colorado Executive director of educator effectiveness, Colorado Department of Education	March 26, 2012	Ms. Anthes provided information regarding Colorado's evaluation reform process. She answered questions regarding Colorado's pilot, cost, and lessons learned.
Amber Arellano	Michigan Executive director, The Education Trust – Midwest	Ongoing	Ms. Arellano has provided ongoing support and guidance by conducting research, building understanding of other states' evaluation systems, and aiding in the navigation of political environments.
Harold Doran	Principal research scientist, American Institutes for Research (AIR)	September 21, 2012	Mr. Doran met with a council subgroup to discuss AIR's experience working with educator evaluation and their value-added services.
Philip R. Fletcher	Senior research scientist, Pearson Education	September 24, 2012	Mr. Fletcher met with a council subgroup to discuss Pearson's experience working with educator evaluation and their value-added services.
Emily Houk	Michigan Vice president, Public Sector Consultants	April and May, 2013	Ms. Houk spoke with a council subgroup to discuss the infrastructure necessary to implement the council's recommendations.
Drew Jacobs	Michigan Data and policy analyst, The Education Trust Midwest	Ongoing	Mr. Jacobs has provided insight into the waiver process, evaluation tools, and other states' reform processes.
Wayne Kuipers	Michigan Government relations, State of Michigan at SAS	September 20, 2012	Mr. Kuipers met with a council subgroup to discuss SAS's experience working with educator evaluation and their value-added services.

Name	Position	Date Consulted	Information Provided
Daniel Lambole	Illinois Principal, Parkside Junior High School	October 12, 2012	Mr. Lambole provided a council subgroup with information about the Illinois State Model for Principal Evaluation and its implementation in his state.
Mariann Lemke	Principal research analyst, American Institutes for Research (AIR)	September 21, 2012	Ms. Lemke met with a council subgroup to discuss AIR's experience working with educator evaluation and their value-added services.
Sarah Lenhoff	Michigan Assistant director of policy and research, The Education Trust – Midwest	Ongoing	Ms. Lenhoff has helped the council understand more about pilots, evaluation tools (particularly student growth tools), and building capacity around evaluation systems.
Chelsea Maupin	Michigan Research and policy coordinator, Pace & Partners	Ongoing	Ms. Maupin has assisted the council with communications around the release of the final report and recommendation implementation.
Sean McLaughlin	Wisconsin Production manager, Value-Added Research Center (VARC), University of Wisconsin-Madison	September 27, 2012	Mr. McLaughlin met with a council subgroup to discuss VARC's experience working with educator evaluation and their value-added services.
Robert Meyer	Wisconsin Director, Value-Added Research Center (VARC), University of Wisconsin-Madison	September 27, 2012	Mr. Meyer met with a council subgroup to discuss VARC's experience working with educator evaluation and their value-added services.
Daniel Murphy	Research scientist, Pearson	September 24, 2012	Mr. Murphy met with a council subgroup to discuss Pearson's experience working with educator evaluation and their value-added services.

Name	Position	Date Consulted	Information Provided
Robert Murphy	New Jersey Principal, East Brunswick High School	March 2012	Mr. Murphy discussed the observation tool that New Jersey currently uses to assess teachers. He addressed the cost, feasibility, and feedback from teachers for the tool.
Mike Nowlin	Michigan Senior public relations and policy manager, Pace & Partners	Ongoing	Mr. Nowlin has assisted the council with communications around the release of the final report and recommendation implementation.
Andrew Rice	Wisconsin Associate director for technical projects, Value-Added Research Center (VARC), University of Wisconsin-Madison	September 27, 2012	Mr. Rice met with a council subgroup to discuss VARC's experience working with educator evaluation and their value-added services.
Julia Simmerer	Ohio Director, Office of Educator Effectiveness, Ohio Department of Education	April 4, 2012	Ms. Simmerer provided information regarding Ohio's observation tools, their training on these tools, and their pilot. She provided insight on the resources that Ohio needs in order for this process to be implemented effectively.
Matt Smith	Colorado Chair, Colorado State Council for Educator Effectiveness, and vice president, Engineering & IT Systems, United Launch Alliance	April 2012	Mr. Smith discussed how Colorado used information from their pilot program to aid the state, administrators, and teachers in understanding and adapting evaluation systems.
John White	Director, SAS Education Value-Added Assessment System for K-12	September 20, 2012	Mr. White met with a council subgroup to discuss SAS EVAAS's experience working with educator evaluation and their value-added services.

Name	Position	Date Consulted	Information Provided
Walter (Denny) Way	Senior vice president, Measurement Services, Assessment and Information, Pearson	September 24, 2012	Mr. Way met with a council subgroup to discuss Pearson's experience working with educator evaluation and their value-added services.
Roni Rucker Waters	Michigan Account manager, Pace & Partners	Ongoing	Ms. Waters has assisted the council with communications around the release of their final report and recommendation implementation.
Nadja Young	Education specialist, SAS Education Value-Added Assessment System for K-12	September 20, 2012	Ms. Young met with a council subgroup to discuss SAS EVAAS's experience working with educator evaluation and their value-added services.

Appendix E: Research and Resources

Other States' Reports		
Document Title	Publishing Organization	Description of Document and Web Link
The State Council for Educator Effectiveness Report and Recommendations (2011)	Colorado's State Council for Educator Effectiveness	This report details the evaluation recommendations made by Colorado's State Council for Educator Effectiveness. Colorado Report
Teacher and Principal Evaluation Pilot Report to the Legislature (2011)	State of Washington's Office of the Superintendent of Public Instruction	This report to the Washington legislature details the teacher and educator evaluation reform process and pilot. Washington Report
The Rhode Island Model: Teacher Evaluation and Support System (2013)	Rhode Island Board of Regents	This guide explains Rhode Island's teacher and administrator evaluation process. Rhode Island Report
RISE Evaluator and Teacher Handbook 1.0 (2011)	Indiana Department of Education, RISE Evaluation and Development System	This handbook details Indiana's teacher evaluation system. Indiana Report 1.0
RISE Evaluator and Teacher Handbook 2.0 (2012)	Indiana Department of Education, RISE Evaluation and Development System	This handbook details Indiana's updated teacher evaluation system. Indiana Report 2.0
Building a Breakthrough Framework for Educator Evaluation in the Commonwealth (2011)	Massachusetts Task Force on the Evaluation of Teachers and Administrators	This framework details the educator evaluation system in Massachusetts. Massachusetts Report
North Carolina Teacher Evaluation Process	Public Schools of North Carolina, State Board of Educations, Department of Public Instruction	This report explains North Carolina's teacher evaluation process. North Carolina Report

Other States' Reports (continued)		
Document Title	Publishing Organization	Description of Document and Web Link
Databases on State Teacher and Principal Evaluation Policies	American Institutes for Research	<p>This website compares the evaluation systems of states across the country.</p> <p>State Database Comparison</p>

Michigan Department of Education Documents		
Document Title	Publishing Organization	Web Link
Professional Standards for Michigan Teachers	Michigan Department of Education	PSMT Report
Michigan's Teaching for Learning Framework	Michigan Department of Education	TFL Framework
Michigan's School Improvement Framework	Michigan Department of Education	SI Framework

Research Papers and Other Reports		
Document Title	Publishing Organization	Description of Document and Web Link
Gathering Feedback for Teaching (2012)	Bill and Melinda Gates Foundation, Measures of Effective Teaching (MET) Project	<p>This report presents an in-depth discussion of the analytical methods and findings from the Measures of Effective Teaching (MET) project's analysis of classroom observations.</p> <p>Feedback for Teaching Brief</p>
Measuring What Matters (2011)	Aaron M. Pallas, Phi Delta Kappan	<p>This paper argues that all states should adopt a new system of program accountability guided by recommended principles.</p> <p>Measuring What Matters</p>
Teacher Evaluation in Michigan (2012)	The Education Trust – Midwest	<p>This report describes Michigan's teacher evaluation legislation and reform process.</p> <p>Teacher Evaluation in Michigan</p>

Observation Tool Frameworks and Resources		
Document Title	Publishing Organization	Web Link
Charlotte Danielson's Framework for Teaching (2011)	The Danielson Group	FFT 2011 Revised
Scales and Evidences for the Marzano Teacher Evaluation Model (2013)	Learning Sciences Marzano Center	Marzano Teacher Evaluation Model
CLASS Implementation Guide (2009)	Classroom Assessment Scoring System, Teachstone Inc.	Class Implementation Guide
5D+ Teacher Evaluation Rubric (2012)	Center for Educational Leadership, University of Washington College of Education	5D Framework
Understand the Teacher Advancement Program	Teacher Advancement Program Foundation	TAP Overview
The Thoughtful Classroom Teacher Effectiveness Rubric: Administrator's Observation Guide	The Thoughtful Classroom	The Thoughtful Classroom Framework Guide
Rating a Teacher Observation Tool	The New Teacher Project	<p>This PowerPoint specifies ways to ensure classroom observations are focused and rigorous.</p> <p>Rating a Teacher Observation Tool</p>

Student Growth Resources		
Document Title	Publishing Organization	Description of Document and Web Link
Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models (2005)	Henry I. Braun, ETS	<p>This paper serves as a review of the opportunities and constraints of value-added models as applied to teacher evaluation. The author argues that value-added models are helpful in identifying teachers in need of professional development and low performing schools, but also includes cautions surrounding technical limitations.</p> <p>Using Student Progress to Evaluate Teachers</p>
Passing Muster: Evaluating Teacher Evaluation Systems (2011)	Brown Center on Education Policy at Brookings	<p>This article provides an overview for evaluating the technical characteristics of teacher evaluation systems and includes worked examples.</p> <p>Passing Muster</p>
The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood (2011)	Raj Chetty, John N. Friedman, and Johan E. Rockoff	<p>This report addresses the long-term impacts of teachers, and viewing those impacts through student outcome data.</p> <p>Long-Term Impacts of Teachers</p>
Evaluating Teacher Evaluation: Popular Modes of Evaluating Teachers are Fraught with Inaccuracies and Inconsistencies, but the Field has Identified Better Approaches (2012)	<p>Audrey Amrein-Beardsley, Linda Darling-Hammond, Edward Haertel, and Jesse Rothstein</p> <p>Phi Delta Kappan</p>	<p>This article argues that many modes of evaluating teachers are not as reliable as their promoters claim, but other options are available.</p> <p>Evaluating Teacher Evaluation</p>
The Colorado Growth Model: Using Norm- and Criterion-Referenced Growth Calculations to Ensure that All Students are Held to High Academic Standards (2011)	William J. Bonk, Ph.D., Colorado Department of Education	<p>This brief paper provides an overview of Colorado's student growth model.</p> <p>Colorado Growth Model</p>

Student Growth Resources (continued)		
Document Title	Publishing Organization	Description of Document and Web Link
A Measured Approach	Daniel Koretz	<p>This paper offers an accessible introduction to measurement issues related to teacher evaluation and value-added models.</p> <p>A Measured Approach</p>
Getting Value Out of Value-Added: Report of a Workshop (2010)	Henry Braun, Naomi Chudowsky, and Judith Koenig The National Academies	<p>This document summarizes the perspective of participants in a 2008 National Research Council workshop on value-added models.</p> <p>Report of A Workshop</p>
Using Student Performance to Evaluate Teachers (2011)	Rand Education	<p>This document summarizes the importance of incorporating multiple measures of teacher performance in an evaluation system.</p> <p>Student Performance to Evaluate Teachers</p>

Non-Tested Subject Resources		
Document Title	Publishing Organization	Description of Document and Web Link
Measuring Student Achievement in Non-Tested Grades and Subjects: Approaches, Issues, and Options for DCPS (2011)	District of Columbia Public Schools	<p>This report documents Washington, D.C.'s system of evaluating teachers with non-tested subjects and grades.</p> <p>DC Non-Tested Grades and Subjects</p>
Measuring Growth for Non-Tested Subjects and Grades (2011)	Tennessee First to the Top	<p>This report documents Tennessee's system of evaluating teachers with non-tested subjects and grades.</p> <p>Tennessee Non-Tested Grades and Subjects</p>

Student Learning Objective Resources		
Document/Website Title	Organization	Web Link
Guide for the Development of Student Learning Objectives (2009)	Austin, Texas ISD – REACH	Austin, TX Developing SLOs
REACH SLO Manual (2011)	Austin, Texas ISD – REACH	Austin, TX SLO Manual
2011-2012 Student Learning Objectives Examples (2009-2010)	Austin, Texas ISD – REACH	Austin, TX SLO Examples
Student Learning Objectives website (2013)	Austin, Texas ISD – REACH	Austin, TX SLO Website
Student Learning Objectives Handbook Version 2.0	Indiana Department of Education – RISE	Indiana SLO Handbook
A Guide to Using SLOs as a Locally-Determined Measure of Student Growth	Ohio Department of Education	Ohio SLO Guide
Indicators of a Strong SLO (2013)	Rhode Island Department of Education	Rhode Island Indicators of a Strong SLO
Guide for Educators: Writing Student Learning Objectives (2013)	Rhode Island Department of Education	Rhode Island SLO Writing Guide
SLO Quality Check Tool (2013)	Rhode Island Department of Education	Rhode Island Quality Check Tool
SLO Examples by Grade/Subject (2013)	Rhode Island Department of Education	Rhode Island SLO Examples

Appendix F: Administrator Evaluation Tools

Administrator Evaluation Tools		
Document Title	Publishing Organization	Web Link
Illinois State Board of Education Performance Evaluation Advisory Council State Model for Principal Evaluation (2012)	The Illinois Performance Evaluation Advisory Council (PEAC)	Illinois Principal Evaluation Model
RISE Evaluator and Principal Handbook (2012)	Indiana Department of Education	Indiana Principal Evaluation Model
Marzano School Leadership Evaluation Model (2012)	Learning Sciences Marzano Center for Teacher and Leadership Evaluation	Marzano School Leadership Evaluation Model
Reeves' Leadership Performance Matrix (2011)	The Leadership and Learning Center	Reeves Leadership Performance Matrix
School ADvance: Administrator Evaluation System for Learning, Growth, and Adaptation	Michigan Association of School Administrators School ADvance Administrator Evaluation	MASA School ADvance Documents and Information

Appendix G: ISR's Administrator Evaluation Tool Structured Review Report

Michigan Council for Educator Effectiveness Administrator Evaluation Systems Focus Groups Final Report

Background

The Michigan Council for Educator Effectiveness (MCEE) was established in June of 2011 as part of Michigan's teacher tenure reform efforts (PA 102 of 2011). The MCEE is charged with developing a fair, transparent and feasible evaluation system for teachers and schools administrators, with the overall goal of enhancing instruction, improving student achievement, and supporting ongoing professional learning.

The University of Michigan's Survey Research Operations Group (SRO) of the Survey Research Center was engaged by the MCEE to conduct ten focus groups across the State of Michigan. The purpose of the focus groups was to gather qualitative data regarding five administrator evaluation tools selected by the MCEE. Participants in the focus groups would be superintendents or individuals in other roles with responsibility for evaluating school principals. The five tools that were reviewed are listed below:

- Illinois State Model for Principal Evaluation
- Indiana (RISE) Evaluation and Development System
- Marzano School Leadership Evaluation Model
- MASA School ADvance Administrator Evaluation System
- Reeves Multidimensional Leadership Performance Matrix

Recruitment

ISD/RESA Recruitment. One intermediate school district (ISD) or regional educational service agency (RESA) was selected for participation from each of Michigan's ten statewide educational regions. In addition, school districts participating in the MCEE Pilot program were to be included in recruitment for the focus groups. The ten ISDs/RESAs selected for the project are listed below, along with the dates on which the focus groups were held:

- | | |
|-------------------------|-------------------|
| • Genesee ISD | February 21, 2013 |
| • Marquette/Alger RESA | February 22, 2013 |
| • Traverse Bay Area ISD | February 26, 2013 |
| • Saginaw ISD | February 28, 2013 |
| • Ingham ISD | February 28, 2013 |
| • Wayne County RESA | March 1, 2013 |
| • Washtenaw ISD | March 4, 2013 |
| • Kalamazoo RESA | March 5, 2013 |
| • Kent ISD | March 5, 2013 |
| • Oakland Schools | March 6, 2013 |

Each ISD/RESA superintendent was contacted by email, with a follow-up phone call from the project manager at SRO. The project manager worked with the ISDs/RESAs to schedule the focus group at a date and time convenient to the ISDs/RESAs and to the overall project schedule. Focus groups were scheduled to allow both travel time to and from the ISDs/RESAs, and some modest schedule leeway in case of inclement weather. ISD/RESA conference facilities were fully scheduled for several weeks in advance, which led to focus groups being held at later dates than were originally desired.

Participant Recruitment. Each ISD/RESA liaison was asked to provide a list of district superintendents and suggestions for charter school representation, from which SRO would conduct recruitment activities. Five ISDs/RESAs chose to contact and recruit focus group participants, rather than having SRO conduct recruitment activities. Those ISDs believed that they would have greater success in obtaining

representation from their membership. The five ISDs/RESAs conducting recruitment were Genesee, Traverse Bay, Ingham, Wayne and Washtenaw. SRO provided guidance about recruitment, requesting inclusion of Pilot and some charter schools from within the ISD/RESA. They were asked, if possible, to attempt to provide a range of rural and urban or suburban, and large, medium and small district representation. They were also asked to avoid supervisor/supervisee pairs in order to allow unconstrained discussion.

In those ISDs where SRO conducted participant recruitment, SRO selected district superintendents and charter school directors or superintendents from lists provided by the ISDs/RESAs. In small districts, all public school superintendents and superintendents/directors from larger charter schools were included in recruitment. In large ISDs/RESAs, prospective participants were selected to provide representation from various-sized districts, rural, urban and suburban areas, and charter schools.

All school districts participating in the MCEE Pilot program were included in focus group recruitment. ISDs/RESAs conducting their own recruitment were given Pilot school information and requested to include the Pilot schools in recruitment efforts. All remaining Pilot schools were invited to a focus group in the ISD/RESA closest to each school's location.

Prospective participants were sent an email invitation to a focus group. Email invitations were individually sent to prospective participants in order to avoid spam filters. The invitation encouraged the superintendent of a district or charter to nominate another person from the school system if he/she were unable to participate. Those participants who did not respond to the email invitation within one week received follow-up telephone contact from a trained survey specialist at SRO.

The response rate to email invitations and telephone follow-up was low, and considerable effort was expended to attempt to contact and recruit participants. When possible, administrative assistants in superintendents' offices were contacted to assist with contact and/or scheduling. When contact was made with someone in the superintendent's office, the district was encouraged to send another representative if the superintendent was unable to attend. Multiple calls were made to non-respondents using scripted messages for answering machines. The project manager at SRO contacted ISD/RESA superintendents to request assistance with recruitment. ISD/RESA superintendents were given a copy of the email invitation, and asked if they could stress the importance of the opportunity to provide input in order to encourage participation by their member-superintendents. These contacts, along with increased telephone follow-up, assisted with boosting attendance at the focus groups.

Participant Mailing. Confirmed focus group participants received an advance mailing timed to arrive approximately one week prior to the focus group. The mailing consisted of a small binder with a cover letter explaining the background and scope, date and time of the focus group; a list of questions to consider as they read through the materials; documentation regarding an administrator evaluation tool (provided by the MCEE project manager); and the State of Michigan Standards for Preparation of Principals. Prospective participants were asked to read the materials in the binder prior to attending the focus group, and to bring the binder to the focus group. In smaller ISDs/RESAs where facilities could accommodate 100% attendance, all non-responders received the advance mailing in addition to those who responded affirmatively.

Focus Group Logistics and Moderation

Focus groups were conducted at ISD conference facilities. Although offered, no ISDs/RESAs accepted reimbursement for use of the facilities. Light refreshments were provided in all focus groups. All participants were given an incentive payment of \$40 cash prior to the start of the focus group. Some participants refused the incentive, and many participants indicated that the incentive would be given to the school district. Consent forms were not required for this project; focus group participants were assured of the confidentiality of the discussion during introductory statements.

Focus group size ranged from 6 to 15 participants, with an average of 9 participants in each session. Seven out of the original 14 MCEE Pilot districts sent representatives to a focus group. Pilot districts were

represented in five out of ten focus groups. Four charter school systems were represented across two sessions.

Focus Group Site	Total Participants at Site	Pilot Districts Represented	Charter Schools Represented
Genesee:	15	2	0
Marquette:	6	0	0
Traverse:	8	0	0
Saginaw:	11	1	0
Ingham:	8	0	2
Wayne:	9	1	0
Washtenaw:	8	0	0
Kalamazoo:	9	2	2
Kent:	8	0	0
Oakland:	10	1	0
Total Participants:	92	7	4

Two SRO project managers conducted all ten focus groups. One person moderated while the other took notes. Moderation and note-taking roles were rotated in each focus group. All focus groups were audio recorded to assist with compiling and verifying notes. Each focus group lasted between 90 minutes and two hours.

A standard moderator's guide was used for all focus groups. The moderator's guide focused on the main framework for the tool (domains, indicators and behaviors, rubric); feasibility; and implementation of the tool. A handout that summarized the domains and subdomains and artifacts/evidence for the tool was given to participants at the start of each session.

The intent of each focus group was to gather input on one administrator evaluation tool by discussing the domains, the indicators and behaviors, the rubric, and the feasibility of implementation. Moderation followed a standard format and focused on the major questions that were sent to prospective participants, with probing questions inserted as necessary to clarify statements or direct the conversation. As noted above, participants received a list of questions to consider as they reviewed the documentation about the evaluation tool to be discussed. The questions helped focus the participants as they read through the materials provided. In many cases the discussion of the participants naturally followed the questions provided. In other cases the participants discussed the major topics, although not always in the order in which the questions were presented.

Focus group findings are qualitative in nature and cannot be generalized to any larger group. Participant comments are based on individual perceptions and experiences. Paraphrased quotations are shown in *italicized text*. Specific districts and commenters are not identified to protect confidentiality.

Results

Discussions in the groups were rich, tended to be lively, and sometimes reached a philosophical and conceptual level. As participants debated the merits of the tool in front of them, discussion ranged from the intent and form of the evaluation process to the nature and complexity of the job of a principal, to how to capture that complexity in the language describing the domains and subdomains, to the priorities between the domains, to the purpose of and use of evidence in the evaluation process. All participants expressed gratitude for being asked to participate in the discussion. All appeared to think the discussion was helpful and brought important issues to the forefront.

There was a great deal of diversity among the types of districts and school systems represented in the focus groups. Districts ranged from small rural districts where one person served in the role of superintendent for the district as well as principal for one or more schools, to urban schools where principals might have responsibility for building management and another role or roles such as that of athletic or transportation director; from small charter schools where the director also served as the principal, to large suburban districts where building principals are supported by several assistant principals, each of whom is assigned a specific area of responsibility (discipline, instructional support, etc.). This diversity led to wide range of opinions about the tools that were reviewed and the relative importance of various features in the tools.

Diversity exists in the evaluation systems currently being used in districts across the state. Current systems range from more complex rubric-based systems to simple narratives. A few participants cited the Danielson model and the Marzano model as frameworks for current administrator evaluation tools, although the form of the actual evaluation process varied from solely narrative to rubric-based evaluation supplemented with narrative. Some districts had recently adopted or were in the process of adopting rubric-based evaluation systems based on and building upon teacher evaluation models. Large districts frequently used some form of electronic system for collecting and filing evidence, such as STAGES or Teachscape. Many districts have incorporated goal-setting and self-evaluation into the evaluation process. An emerging practice that was mentioned in several focus groups was the concept that all evaluations in a school or district, from teachers up through administrators and central office staff, would build upon each other and be linked to a central set of district goals. This practice was described as having the evaluations “roll up” from one level to the next. The features most often cited as being “liked” about the current systems were the inclusion of a goal setting exercise and processes that led to an in-depth dialog between the evaluator and the principal. Simplicity and efficiency with respect to the amount of time required to complete the evaluation process, and inclusion of a self-assessment or reflection were also mentioned by participants. Other items mentioned as being “liked” were integration with school or district goals, inclusion of a narrative section and the use of evidence.

It is perhaps understandable, given the diversity of the school systems represented and the various tools currently in use, that there was no general agreement across the focus groups about a “best” model for administrator evaluation. Participants disagreed within and across groups about the features and frameworks of the tools reviewed. For instance, while some participants cited the Marzano system as being a model for administrator evaluations, participants in the two groups that reviewed the Marzano tool disliked its length, its perceived complexity, and its lack of attention to management skills. We can therefore make no recommendation regarding preference among the five tools that were evaluated. Although there was little agreement within or across groups regarding tools, common themes did emerge from the focus group discussions regardless of the tool being discussed. These common themes, presented below, provide important information to be considered when discussing selection of an administrator evaluation tool for the State of Michigan.

Themes

Purpose and Intent of the Evaluation Tool

- All groups recognized that the State desires a summative rating of principals. Participants expressed a desire for a tool that is *formative* as well, and can be used for growth and development of principals. The tool should be supportive of emerging principals as well as providing guidance and feedback for experienced principals.
 - *We want to avoid this being a documentation tool as opposed to a growth and improvement tool.*
 - *We encourage people to stretch professionally and set high goals. We have to make it a safe environment to do that.*
 - *We want a tool that is formative rather than summative.*
 - *We know the State wants the summative number at the end. We want to emphasize the formative piece--that is what we are really trying to accomplish.*
 - *We don't want this investment to be a compliance exercise. This should be more of a process that evokes growth.*

- *Evaluation is an event. Feedback and growth is a process. We want the process that results in growth.*
- The usability of the Michigan Standards for Preparation of Principals was raised as an issue in at least three focus groups. Participants commented that the standards were too long and detailed, and out-of-date. Many participants stated that the standards should be updated, with some agreeing that only after revision should an evaluation tool be selected to support the revised standards.

Communication from the MCEE

- Gathering input from the districts (by conducting focus groups) prior to selecting a tool was viewed very positively by all participants.
 - *Thank you for coming out and hearing from people in the field.*
 - *Thank you for this process.*
 - *Thank the MCEE. We have had so much more feedback and opportunity in this process than any other time.*

Adaptability

- The great diversity among the types and sizes of districts was broadly recognized and discussed. In addition, there was broad recognition that the job of a principal is complex, and that the focus of any particular principal will vary greatly by building and/or district, or by year within the same building or district. The diversity among the districts, and the variety and complexity of the job of the principal, led to the desire to customize or adapt tools to meet local needs and priorities.
 - *One size fits all doesn't work. The evaluation system will need to be adapted to each school district.*
 - *We need to be able to reflect the differences between buildings, districts, size and local issues in the schools. We need that flexibility.*
 - *We need to have flexibility in terms of setting the priorities underneath each domain. My district may be different from your district.*
 - *It must have the flexibility to be an effective evaluation tool for evaluation in a high urban area with lots of social issues, versus a one-room school house in the Keweenaw Peninsula.*
- Adaptation strategies could involve deleting, adding or placing increased emphasis on domains or subdomains, or determining definitions based on local needs and the local school improvement plan. Additional adaptations could involve adding elements such as self-evaluations or narratives to complement the rubric.
 - *Instead of the shotgun approach, this could be customized to a building or to a need.*
 - *Maybe we should be defining some of the rubric based on school improvement plan or district initiatives. Then you get alignment to district on school improvement plan, rather than using an off-the shelf tool.*
 - *All of our principals in our districts have different roles. We need to use the tool a la carte, to choose domains to fit the principal. We need flexibility to apply the domains.*
 - *We need to allow for flexibility to add locally-established goals that are set by the School Board.*

Rubric

- All agreed that a simple but comprehensive model or rubric, based on the most current research on leadership, that could be adapted locally, would be best. In general, rubrics were felt to be helpful because they removed some level of subjectivity from the evaluation. They could also be seen to be too complex or rigid to be easily used in the variety of district sizes and configurations found across the State of Michigan.
 - *Rubrics help the evaluator and the person being evaluated. It is not about who you like.*
 - *I like that it forces us to start in one place to help us make a less subjective judgment.*

- *Rubrics are great for defining expectations and painting a picture of what we want it to look like.*
 - *The rubric style is good, but it needs to capture the range of a principal's performance.*
 - *Principals in different schools and districts have different challenges (climate, financial, expectations from student homes, school readiness, etc.). Superintendents need to have the flexibility to modify the tool to the situation.*
- There is a feeling that long rubrics are not helpful and will potentially damage the dialog between the evaluator and the principal by diverting attention to the details and checklists, and away from growth. Long rubrics with many domains and subdomains are described as overwhelming. The focus on improving performance will become lost in the detail. *"Less is more"* was a phrase stated in several focus groups.
 - *Less is more. Setting two or three goals to focus on some of these concepts would be stronger.*
 - *If you have to go through all of these [domains], the effect gets watered down and it might not move the building forward.*
 - *Quantity isn't quality. There are WAY too many domains [in the rubric].*
 - *I feel overwhelmed. As you look at each one of those things and try to give equal judgment to all of those things, it becomes too much.*
 - *This will become a compliance exercise.*
 - *This will become perfunctory. Get it done, and go through it, because there is so much. It makes it hard to do anything with any validity at all.*
- If the rubric is long and contains many domains/subdomains, evaluators will want to focus on a few key items in each evaluation period (adapting to local priorities and needs). A few participants asked for identification of, or emphasis on, *the highest leverage* skills or important domains identified in the research, in order to narrow the focus and reduce the length of the rubric.
 - *Go through the rubric and pick a goal for people who are doing well.*
 - *Less is more. Setting two or three goals to focus on, using these concepts, would be stronger.*
 - *I would distill these for a particular person and that person's needs. You have to distill it down to a plan that you work on collaboratively with that specific individual.*
 - *That scope is just so huge, it reaches a point that you have to focus on some very specific components.*
 - *I would not evaluate on every single one of the areas; I would focus on the areas where the person is struggling.*
- School board members who conduct evaluations will have trouble understanding and using complicated rubrics.
 - *How would the board do this? They do not understand the vocabulary.*
 - *A user would need underlying knowledge of instructional practices.*
 - *It will take coaching to ensure consistency in how tools are being used by the boards.*
- The presence of a rubric is generally helpful to start evaluation conversations and mediate any legal challenges.
 - *It is the conversation that stems from the rubric that is important.*
 - *The dialog is the most important thing in the evaluation. The instrument shapes the dialog.*
 - *We must have metrics that are reliable and valid. We have to protect ourselves in case of layoffs or personnel issues.*
 - *The end product must be defensible on the witness stand.*

- A majority of participants requested more specific language in rubrics, citing the lack of clarity around descriptors of behavior, and personal viewpoints regarding meaning of certain words or phrases.
 - *What does it (good performance) look like, what does it sound like, what does it feel like?*
 - *The individual definition of the words is a problem. Words like “frequently” mean different things to different people.*
 - *If definitions are subjective, you lose the goal of consistency in implementation across the state.*
 - *Maybe the tool isn’t good enough if we have to spend a lot of time defining what each cell means.*
- Others felt less specific language helped the districts define local needs.
 - *I’m ok with that loose definition. Looseness allows for conversations to start, you define what that means in your district.*
 - *I like this rubric. Other rubrics use words like “always” and “never”. This gives you a wider base on which to make a decision. You can quantify items in a discussion with the principal.*
- Relative weighting of the domains and subdomains in the rubric, and how the relative weights would or should contribute to a summative rating was not clear to many participants. Participants discussed allowing districts to use individual, building or district-level goals to guide the weighting. However there was some recognition in at least one group that prescribing weights removed a level of subjectivity from the evaluation process. When guidance was provided in the tool, the prescribed weighting of domains was sometimes viewed negatively and deemed to be unresponsive to local needs.
 - *How would these domains be weighted? Some have more elements than others—do those domains carry extra weight?*
 - *Weighting (domains) should be flexible – by district – I don’t think there is a one size fits all for weighting it.*
 - *Can local districts weight the domains based on local goals?*
 - *What are the metrics behind these evaluations? How would these categories be weighted?*
- In at least two groups, there was discussion about summing ratings across domains, and what methodology or weighting should be used to arrive at an overall rating. It was felt that those items which would cause the most impact in the buildings or districts (management skills, interpersonal skills) should be weighted more heavily, and should explicitly be included in the rubrics. With the exceptions of management, interpersonal skills, and professional integrity, there was no agreement as to which items should be weighted more heavily to develop a summative rating. It was suggested by some that any weighting should be based on the most current research on school leadership with the goal of improving student learning.
 - *What if you have someone who does well on a rubric, but isn’t a good fit for your district?*
 - *What if the person scores well in most areas, but one area is so bad that it has an adverse impact on the school? Will that one area impact the rubric appropriately?*
 - *How many domains have to be unsatisfactory for it to result in serious action? Could one unsatisfactory be a deal breaker?*
- Summative ratings that differed from the State of Michigan required language were noted. Participants wondered if the rubrics would be valid after modification to match the State of Michigan language.
- Participants in several focus groups expressed concern about the calculation of student growth and achievement. There is concern about reliance on the MEAP, although some districts do not have resources to pay for other tests that might be allowable. Participants in all focus groups

noted that they would like to have a decision on, and information about, the process for calculating student growth and achievement.

- *Other tests might be more important to measure growth through the year than the MEAP.*
- *We need to look at other data (for growth) instead of just the MEAP.*
- *The student growth piece is very grey.*
- A goal setting/self-evaluation component was viewed by many as critical to the evaluation and growth process. Some viewed a narrative component as helpful for providing context to a summative rating. The narrative leads to conversations that are critical to the evaluation process
 - *Without a self-reflection piece it becomes another checklist.*
 - *Self-assessment – that is missing – it starts the conversation.*
 - *Using a self-reflection, focusing on a couple of elements... is a way to get flexibility and turn this into a growth piece.*
 - *If I had to live with this, I would incorporate a self-reflection piece to have the principal rate himself/herself on each of the areas.*
 - *The narrative is the most valuable part of the evaluation.*

Domains and Subdomains

- The complexity of the job of principal led to much discussion and debate about the domains and subdomains in each of the rubrics and whether those domains covered the tasks and skill sets needed by a principal.
 - *The administrator's work is a combination of leadership and management skills; they work on technical issues and adaptive issues; they work on planning and execution. I look to see where these things fit within the domains. The domains should reflect these skill sets. We need to evaluate both leadership and management.*
 - *We require principals and building administrators to be experts in organizational management and be able to distill, create, operate, evaluate.*
 - *I view the principal as instructional leader for students and setting those goals but also as instructional leaders for training teachers, for food service workers, for custodians, and different kinds of needs for families – the principal has an ever-expanding role in building the capacity of all.*
 - *I'm struck by the complexity of the principal role when I see the domains isolated like this.*
- Participants agreed that the ordering of the domains in documentation sends a distinct message as to the importance of the domain. It was not clear from the documentation provided as to why one domain was listed before another in the rubrics. It would be helpful if the ordering and selection of domains were explained to users in order to enhance understanding and acceptance of the tool.
 - *What is put first on the list sends a message. The ordering of the domains should be different.*
 - *Principals have a lot of additional duties. Results shouldn't be #1.*
 - *These domains should be re-arranged.*
- Two domains were repeatedly mentioned as missing from the rubrics: building management and "soft skills". Many tools did not either mention or emphasize building management, or "soft skills" (such as communication, conflict management, problem-solving, or stakeholder communication), to the extent that participants felt was necessary. It was repeatedly stated that along with providing instructional leadership, these were key components of a building manager's job. It was also felt that not having these items specifically called out in an evaluation tool might lead to difficulties in terms of moving someone out of a position. Also potentially missing was recognition of the increasing amount of time that principals spend on evaluating teachers.
 - *This is a wonderful tool for evaluating instructional leadership, but there is a lot more to being a building principal.*

- *If you are struggling in managing your building, in professionalism, in communication, then you are not impacting student learning and you can't be evaluated with this rubric. The rubric doesn't address management.*
- *If you are not a good manager, you will not be successful in creating a learning environment.*
- *You can be book-smart, but if you don't have the personal relationship-building characteristics, you won't be able to lead that school.*
- *There needs to be strong focus on getting along with people.*
- *Teacher evaluations are not emphasized enough.*
- Items currently in the domains that could or should be removed were mentioned.
 - In some districts, principals are viewed as implementers of district policy, curriculum and systems, rather than serving as developers. Tools that include development of district policy or curriculum must be adjusted for local conditions.
 - In many districts, principals do not control teacher recruitment and hiring; in some districts the principal may have input in teacher hiring. Tools that have language to suggest that the principals control or are responsible for teacher hiring (or other personnel actions) must be adjusted for local conditions.
 - Roles such as assistant principal, special education director, athletic director, dean of students, etc. may not have tasks in all of the domains. Rather, they may focus in a specific area such as discipline or attendance. Domains, subdomains and indicators must be adjusted for these roles. Similarly, use of the rubrics for other central office personnel will require adjustments.

Evidence/Artifacts

- Types of evidence or artifacts can be suggested, but should be locally determined or defined to meet the needs of the district. Some tools were not clear as to whether evidence items were mandated or merely suggested. Participants expressed a preference for "suggested" evidence.
 - *I assume this is a brainstormed list and you don't have to do all of these.*
 - *I would have the principal define what evidence is needed. We ask them "what does growth look like in your area."*
 - *Mandating use of particular items is not okay. Let the districts and schools have choices.*
 - *Local districts should be able to define what is required as evidence.*
 - *The organization needs to spend time shaping what the district will consider evidence of reaching those goals... what the district defines as evidence.*
- Longer rubrics with extensive lists of required evidence may risk becoming compliance exercises, rather than focusing on growth and development of administrators. In some cases, participants asked for identification of the "highest impact" evidence items as noted in the research. Some participants felt that required evidence should be determined locally in order to make the process more meaningful to principals.
 - *We don't want people to prepare like this is an IRS audit.*
 - *Presence of evidence does not necessarily demonstrate efficiency or effectiveness.*
 - *Evidence is only useful if the principal is using the evidence to understand the current situation, share that understanding with teachers and staff, and develop and implement remediation plans to move the school toward stated goals and objectives.*
 - *We don't want this to be about gathering evidence; we need to have that discussion about growth.*
- An electronic means of collecting and storing evidence (such as STAGES) is very helpful in large districts.

Implementation and Training

- Participants indicated that no new personnel would be hired in response to the adoption of a new evaluation tool. Rather, the superintendents will need to find time to conduct evaluations

using the new system. Technology may be needed to facilitate the evaluation process, especially in larger districts where one superintendent must evaluate a large number of administrators. As noted below, districts may need assistance with the cost of licensing or technology.

- *The person who is doing evaluation now is the same person who will do evaluation later. They will just have to let other work go to make the time needed to get the work done.*
- *The resources are “us”. We cannot add resources. We will do it, but something else will have to give.*
- Professional development/training will be needed to ensure consistency of application both within districts, within ISDs, and across the State. The reasons for inclusion of each domain and sub-domain, and how that item contributes to improved instructional leadership should be included with the documentation to facilitate understanding and acceptance of the tool. There was considerable concern expressed about establishing inter-rater reliability for consistent application of the rubric and legal defensibility.
- Participants discussed conducting internal training, or holding extensive debriefing sessions with users in the districts to develop a shared understanding of the terminology and expectations. In most focus groups, the ISDs/RESAs were suggested as focal points for provision of training and ongoing expertise before, during, and after implementation. It was also noted that some type of training program or support would be needed for school board members who will use evaluation tools.
 - *It would be worthwhile having a day-long training with the superintendents and principals, sitting in the same training session, to develop a shared understanding.*
 - *There is a rationale to say that everyone has the same basic training from a fairness and legal standpoint, even if it is a webinar or on-line. It would help to provide inter-rater reliability across the ISD/RESA and across the state.*
 - *We would want a local expert, perhaps at the ISD/RESA, who has been trained extensively to provide support. This would provide an ongoing conversation about the tool and a liaison for discussion of problems.*
 - *Training could be under the umbrella of the ISDs/RESAs. The ISDs/RESAs could build a support network.*
- Implementation should not be rushed. Time must be allowed for planning, training and roll-out. A piloting period is needed, with time for feedback and discussion after seeing the tool in action, followed by appropriate adjustments.
 - *Take the time to get it right.*
 - *We cannot implement a new system overnight.*
 - *Do you want it done, or do you want it done well?*
 - *A critical mistake on the State’s end would be to tell us in July that we have to roll this out in August. That will never work.*
 - *Implementation will take time. There needs to be pilot, and it will probably not be perfect. We’ll need time to monitor and adjust no matter what the system.*

Feasibility

- Most participants indicated that their districts would be “good soldiers” and implement any tool selected by the State.
 - *We’re all professionals. We can implement anything.*
 - *We are all good soldiers. We do what we are told.*
 - *Nothing is a showstopper—we can make anything work. But do we want to comply or do we want to do something that is useful and meaningful? In this region we are hungry for something that is useful and meaningful.*
- Funding was raised as an issue in most focus groups. Funding could prove to be a barrier to implementation, especially in districts currently running or projecting budget deficits. Funding

would be needed for licensing/use fees, and any necessary technology (i.e., iPads, online systems). Most participants felt that the State should provide funding or assistance with funding for the costs of training.

- *[This] should not become an unfunded mandate (repeated in several focus groups).*
 - *We can't pay for this. We are a deficit district.*
 - *Cost will be an issue for everyone, everyone is already stretched.*
 - *If mandated, it should be funded; it will take a huge amount of money. The money is not there.*
- Lack of adaptability to local conditions was cited as a barrier to implementation by a few participants. The majority of participants simply assumed that a tool could and would be adapted to accommodate local needs and priorities.
 - *Can we implement this [as given]? Yes. Should we? No.*
 - *We could implement this model and it would be meaningless. Take the key points and let the local districts adapt to their environment and need.*
 - *Lack of flexibility is a show-stopper for me.*
 - Adequate time must be provided for implementation. Many participants noted that a significant amount of time would need to be invested both by district leadership and by principals to understand, adapt and implement a new tool.
 - *I would bring in all of my administrators to discuss what words mean, so we are all on the same page.*
 - *Local districts would need time to create local definitions and understanding.*
 - *Our teaching and training would revolve around developing common understanding of meaning for our district, our community, our population, our buildings.*

**Focus Group Notes
Summary by Tool**

The following documents present a tool-by-tool summary of comments and notes from the focus groups. Notes from the focus groups are grouped by major topic, and discussions have been summarized to show the diversity and range of opinions and comments. Discussions in the group often shifted across topic areas, regardless of the prompt presented; notes are therefore not in chronological order.

Focus group findings are qualitative in nature and cannot be generalized to any larger group. Participant comments are based on individual perceptions and experiences. Paraphrased quotations are shown in *italicized text*. Specific districts and commenters are not identified to protect confidentiality.

Summaries are presented in the following order:

- Illinois State Model for Principal Evaluation
- Indiana (RISE) Evaluation and Development System
- Marzano School Leadership Evaluation Model
- MASA School ADvance Administrator Evaluation System
- Reeves Multidimensional Leadership Performance Matrix

Illinois State Model for Principal Evaluation

Domains

Participants in the first focus group were somewhat more positive about the domains than those in the second focus group. Participants in both groups agreed that the list of domains and subdomains, and the Illinois tool overall, was far too long. Participants described the tool as *quantity instead of quality; overkill and duplicative*. One participant stated that while he liked the domains, the rubric was so large that *you can't get your head around it*. Another participant commented that the large amount of content made it difficult to discern the most important items. There was broad agreement in both groups that 'Improving Teaching and Learning' is the most important domain.

At least three participants in Group 1 stated that they liked the Illinois domains. One participant in Group 1 stated that he could find research on effective schools throughout the rubric. On the other hand, at least two participants in Group 1 felt the definitions of the domains and subdomains were unclear and subjective. *The definitions lack understanding. Some of these things are subjective.* Another felt, with some support from the group, that data-driven decision-making was not covered adequately, and should be called out in its own domain. One participant liked that 'Collaboration' and 'Culture of High Expectations' were presented in separate domains. One participant in Group 1 suggested the rubric could be built around 'Improving Teaching and Learning', with all other domains leading to and supporting Improving 'Teaching and Learning' in order to reduce the perceived number of domains and make the rubric appear more manageable. Group 1 discussed possible domains to eliminate, but no consensus was reached. All felt that the domains which included integrity, school culture, and communication issues were critically important and could not be eliminated.

Participants in Group 2 did not agree on the content of the domains or the tool. Some participants liked the tool and the domains, others did not. There was consensus in Group 2 that the domains and categories were vague, and would be difficult to evaluate. There was also consensus in Group 2 that the tool did not include enough emphasis on communication, collaboration or problem-solving skills. Some participants felt that leadership skills and behaviors should receive more emphasis; others felt that planning behaviors should be more specifically mentioned in the domains and subdomains. At least two participants disagreed with the definition of the 'Leading and Managing Systems Change' domain, but each for different reasons. One participant felt that the description was fine, but that decisions and activities shown in this domain would be done at the district level. The other person disagreed with the definition of the activities listed under 'Systems Change'. The participant felt that 'Systems Change' should reflect an engagement of the staff in order to change the culture and direction of the school.

Participants in Group 2 discussed the complexity of the job of a principal and the diversity of situations in districts across the state. *One size does not fit all for Michigan districts* was stated by one participant, with agreement from many in the group. Group 2 participants discussed possible re-arrangement of the domains to more clearly emphasize that communication and collaboration were required elements in each domain, or that all of the items in the domains are done in support of 'Improving Teaching and Learning'. One participant felt that student achievement and growth should be shown in its own domain; another requested more emphasis on teacher evaluations.

There was consensus in both focus groups that flexibility would be needed to modify the tool for various roles, and building and district goals or needs. Similar to participants in Group 1, the Group 2 participants viewed the role of principal as that of an implementer, rather than a developer, and noted that some language in the domains would need to be changed to reflect implementation activities.

Both Groups briefly discussed the relative importance of domains. While 'Improving Teaching and Learning' was viewed as the most important domain, infractions in the 'Leading with Integrity and Professionalism' domain could lead to a principal being terminated. In Group 2, it was agreed that 'Improving Teaching' should be listed at the first item in the rubric, however the 'Integrity' domain may ultimately carry more importance depending upon the situation. Participants in Group 1 did not come to

any conclusion about ordering of the domains, but agreed that the 'Integrity' domain was critically important.

Indicators and Behaviors

Participants in both groups noted that the Illinois rating terminology was inconsistent with the wording required by the State of Michigan. One participant in Group 1 felt that the Illinois rubric did not mesh well with the State of Michigan Standards for Preparation of Principals.

There was consensus in Group 1 that different indicators would be needed for different administrative roles. It was noted that the indicators needed to be revised to recognize the difference between high schools and elementary school administrators. For instance, high school completion rates and college admissions currently included in the indicators would not be appropriate for an elementary principal. There was broad agreement that districts would need flexibility to modify the tool for role, buildings and district-level focus. One participant remarked on an indicator that seemed to say that it would be okay to be non-compliant with FERPA; *you can't be "basic" on FERPA or you will be breaking the law*. At least two participants endorsed adding indicators to the 'Integrity' domain to demonstrate respect for all backgrounds and cultures.

Participants in Group 2 felt that the expectations for performance in the categories were not clear, and descriptors needed to be more concrete. The verbs and descriptors used in the indicators could mean different things to different people, and the behaviors were not clearly laid out. *We would need to unpack this to develop a shared understanding* stated one participant in Group 2. Another participant stated that it was hard to distinguish between the Basic and Proficient categories. Group 2 participants felt that it would be difficult to establish inter-rater reliability due to the unclear language in the rubric, allowing the potential for many interpretations to arise.

Participants in Group 2 mentioned a number of items that they felt were not well covered in the indicators: collaboration and communication, leadership skills, and planning and implementation skills.

Evidence

At least two participants in Group 1 liked the list of suggested evidence, and one person noted that the suggested evidence was supported by research. The participants felt that having a common presentation of suggested evidence would help with conversations. It was felt that the evidence should not be the focus of the conversation or the evaluation. *Evidence is only as valuable as the principal's ability to use it for moving the school toward goals* stated one participant. The participants in Group 1 agreed that mandating use of specific examples would not be acceptable; participants felt that districts must have choices in terms of defining evidence of district or building goals and initiatives.

Group 2 was far less enthusiastic about the list of suggested evidence. Comments included: *This is a pile of minutia; this will become a laundry list for compliance; some would like to see this as a checklist*. One participant noted that some of the items listed for evidence are controlled by the district, and are not appropriate to include in an evaluation of a principal. Other evidence items were thought to be more subjective in nature, such as the quality of newsletters. Participants in Group 2 also reacted to the length of the evidence list and felt that list could be reduced by focusing on those items that were supported by research, and therefore more important. One participant noted that there was a *danger* of the list turning into a compliance exercise for principals. At least one participant disagreed, viewing the list as *opportunities, instead of prescriptions*.

Participants in both groups generally agreed that the districts should have flexibility to define evidence according to building or district needs. Most agreed that particular items of evidence could be suggested to districts, but not required. *The evaluation is an event. Feedback and growth is a process. We want the process that results in growth. Don't make this into an audit event by requiring all of this evidence. We don't want people to prepare like it is an IRS audit*. Participants in both groups felt that it was important to have the principals to show how they are using data to identify a problem, then develop and implement a plan to remedy the problem.

Rubric

Group 1 participants were more positive about the rubric, with the caveat that flexibility would be needed to meet various needs across roles and districts. There was agreement that the tool was too long and could be overwhelming, leading to the desire to narrow the focus to a few key areas. Comments included: *I like the tool, and would welcome the opportunity to use this with my administrators; It feels user-friendly, and is not filled with jargon; I feel that these are really well written.* At least one participant in Group 1 felt that the rubric provided adequate guidance to move from one level to the next. Participants agreed that it was the conversation that stems from the rubric that is really important, rather than the rubric itself, however the rubric must facilitate conversation.

Group 2 was far less positive about the rubric, repeatedly citing unclear verbs and descriptors. Comments from participants included: *This is not a clear document-I can't tell what is needed from reading this; Everyone would define the wording differently; The behaviors need to be clearly defined, and the evidence needs to support the behaviors; Maybe the tool isn't good enough if we have to spend a lot of time defining what each cell means.* The group discussed the difficulty of achieving fidelity and inter-rater reliability within and across districts, and the potential for lawsuits without better definitions in the rubric.

Group 2 participants also discussed the length of the rubric. Many participants felt that the rubric was too long and that the list of domains should be reduced. *Less is more* was stated several times during the conversation. Several Group 2 participants stated that they would choose to focus on a few domains and subdomains each year, rather than tackling the entire rubric. At least two participants suggested that the rubric be compared to the most current research on best practices to differentiate the most important domains and to remove less-important items.

There was consensus in both focus groups that flexibility and local adaptability would be necessary if the districts were to attempt to use the tool: *One size does not fit all.* One participant stated that he would prefer a tool where teacher and administrator evaluations could be linked, and would build on and use student achievement data.

Feasibility/Implementation/Training

Some participants in Group 1 were more positive about feasibility; others felt the tool was too long, and stated that fewer domains would be better. All participants in the group agreed that a significant investment in professional development and training would be needed to successfully implement the tool. Similarly, all participants in Group 1 felt that flexibility and adaptability in use of the tool would be critical to implementation. Flexibility was mentioned multiple times during the conversation. Group 1 Participants in Group 1 discussed the use of technology. Those participants currently using an electronic documentation management tool felt that the use of such technology could help to accelerate implementation. At least one participant stated that the tool must be on-line to facilitate implementation.

Group 2 participants did not view the Illinois tool as favorably. At least one participant liked the portfolio of evidence, however the majority in the group were not in favor of using the rubric. There was agreement in the group that the tool would be difficult to implement due to vague language in the rubric. While participants in Group 2 agreed that the Illinois system could be implemented, participants agreed that it would not be their tool of preference: *Could we implement this tool? Yes. Should we? No.*

Participants in both groups agreed that extensive training would be required prior to implementation. Training would be needed to understand the tool, to develop a common language, and to develop inter-rater reliability. *Implementation without that investment would backfire.* Participants in both groups worried about the process for certification of evaluators, which was called *lengthy and extensive*. There was concern that some districts would lack a certified evaluator. One participant suggested adopting and adapting the tool for superintendents and central office workers to increase the efficiency of training and reduce the amount of training investment for the district overall. Participants in Group 2 discussed and agreed that partnering with the ISD/RESA for training and evaluator certification would be helpful.

Participants in both focus groups agreed that domains and components would need to be tailored to the specific position (assistant principals with limited focus, athletic or special education directors, deans,

etc.), or to the building or district plans. If flexibility in use of the tool were allowed, one Group 1 participant stated that he would align all positions in the school to the same rubric. Many participants in Group 1 agreed that they would use *focus areas* to tailor the application to a principal, building or district.

Participants in both groups agreed that a pilot would be needed prior to implementation to allow the users to become familiar with the language and use of the tool. The pilot should cover rural, urban, and suburban schools. There was agreement in Group 1 that the State should take time to carefully consider the selection of an evaluation tool: *take your time, get it right*. Several participants stated that cost would be an issue for their districts. Cost items included licensing, training, and technology for implementation.

Other

Participants in both focus groups felt strongly about finding a tool that promoted meaningful conversations with principals about growth and development: *Evaluation is a process, not an event; We want the process that results in growth*. Participants in both groups expressed concern over relying solely on the MEAP to measure student achievement, and encouraged the inclusion of other data sources. Other tools suggested by participants included the MASA School ADvance System, the Danielson Systems, the 5-D System, and the Massachusetts System.

The Michigan Standards for the Preparation of Principals were discussed in both focus groups. Participants in both focus groups stated that the Standards were out-of-date and need to be updated. One participant in the first group stated that the Standards attempted to provide too much detail, and lacked a focus on instructional leadership. A participant in the second group stated that the Standards should be updated before an evaluation tool is chosen for the State. There was broad agreement in both groups that the Standards are not useful in their current form.

Indiana RISE Evaluation and Development System

Domains

The participants in the first focus group commented extensively that the Indiana RISE System was focused on instructional leadership, and did not include key management tasks or skills in the list of domains: *The day-to-day work of the administrator is not in here; The domains do not fit.* Lack of attention to these key management responsibilities was noted as a critical deficiency in the Indiana RISE system. At least one participant in Group 1 felt that the tool did not provide sufficient coverage of building climate and culture, and another felt that administrator professional development was missing.

One participant in Group 1 commented on the use of ‘Student Learning Objectives’ in the domains. The participant felt that it did not make sense to have a set of goals or objectives in the evaluation system that differed from the school improvement plan; others in the group agreed. One participant felt that the administrator evaluation should be linked to teacher evaluations, with both teacher and principal evaluations linked to the school improvement plan. Other participants in the group noted that the Indiana Rise System did not reflect the Michigan Standards for Preparation of Principals. Another participant also noted that the Michigan Standards for Preparation of Principals were very outdated.

Group 2 participants also noted that management was not covered in the domains, however that was not the focus of the discussion. There was discussion of the importance of communication, culture and relationship building skills for a principal. While many people in Group 2 agreed that communication, culture and relationships were not represented well-enough in the Indiana RISE domains, another person suggested that the Human Capital domain could be re-named or re-tooled to cover those skill sets. One participant felt that the RISE System did not sufficiently capture innovation. However, another found items that linked to innovation embedded elsewhere in the domains and subdomains. One participant noted that the tool lacked a domain related to implementation of a viable curriculum.

Some participants in Group 2 questioned the prioritization or ordering of the domains. One person in the group noted that the tool did not follow Michigan law regarding evaluations, which allows inclusion of other facets of school performance rather than focusing on test scores: *Are we going to change the Michigan law to meet this tool?* More than one person in Group 2 noted that the RISE System focuses on the summative performance of students, rather than the formative performance of students (and therefore principals): *I have trouble focusing on just the summative performance.*

Participants in both focus groups stressed that the domains and sub-domains in the RISE tool would need to be changed or adapted to the different roles and responsibilities found across districts. Responsibility for hiring teachers was a task that participants in both groups called out as being done at the district-level, and outside of the control of the principal. Participants in both groups felt that the tool should incorporate, or be adapted to, the local school improvement plans: *All evaluations must be directly linked to the school improvement plan—otherwise it is not meaningful to people.* One participant in Group 1 spoke of needing to use the tool *a la carte* in order to adapt to local roles and responsibilities or other conditions.

Participants in both groups expressed considerable concern about the length of the list of domains and subdomains in the RISE tool: *I’m interested in a robust tool where less is more—I’m not sure that RISE is it.* Participants in both groups expressed concern about the number of indicators and the length of time it would take for a principal to complete an evaluation: *This will become perfunctory—get it done, get through it, because there is so much; Do you want it to be meaningful, or do you want us to get it done?; This seems like a checklist.* The participants suggested that the application of the tool could be improved by focusing on a few key areas or goals for the evaluation period; *If your time is short, zero in on the problem; Go through the rubric and pick a goal for people who are doing well.*

Indicators and Behaviors

In general, participants in Group 1 expressed some dissatisfaction with indicators listed in the Indiana RISE tool. In some areas the language was felt to be too subjective, in others it was not subjective

enough. One participant in Group 1 expressed concern that the semantics in the tool would not work across the various districts; districts may wish to define items differently. Another participant felt that the language in the 'School Achievement' domain allowed too much subjectivity, and might allow a principal to set a less-than-rigorous goal. At least two participants in Group 1 expressed discomfort with the language in the 'Professional Development' domain. One participant felt that the 'Professional Development' domain had weak language and was poorly defined. The participant felt that the language "Provides opportunity for..." could become a way to deflect responsibility for engaging the staff (as in "I provided the opportunity, but they did not take it"). Another participant noted that the language in the 'Professional Development' domain addressed only teachers and did not address the administrator's professional development.

The lack of attention to building management in the indicators was noted as a serious deficiency in Group 1. At least two participants in Group 1 felt the indicators and rubric should be cross-referenced to effective schools research, or to the Marzano literature. The group discussed flexibility in general, and the need to modify the tool for various role, building and district needs. Most discussion of indicators (and the rubric in general) was couched in terms of needing flexibility, and needing to incorporate school improvement plan goals: *We need flexibility to apply the domains--one size does not fit all.*

Group 2 felt strongly that the indicators were not consistent across the performance ratings. There was broad agreement that the Level 4 (Highly Effective) language was structured completely differently from the other levels, and appears as if it was added to the rubric at a different time. Group 2 also noted areas where flexibility would be needed to adapt the tool to role, building and district needs.

While one person in Group 2 felt it would be easy to make a decision based on the language in the indicators, many others in the group felt that the subjective language in the indicators would make it difficult to make a rating decision: *I think it would be difficult; The ambiguity or subjective nature of this makes it hard—what am I measuring?* Participants in both groups commented on the number of items listed in some of the boxes. It was felt that making a decision for boxes with a large number of items would be more difficult than if there were fewer items in a box.

Evidence

Overall, participants in Group 1 liked the list of suggested evidence. Comments from Group 1 participants included: *it is a great compilation of data; It is pretty good.* Most participants in the group interpreted the evidence list to be suggested as opposed to mandated. Many participants made suggestions for items to add to the list, such as budget documents, 360 degree feedback, surveys, and communication items such as newsletters and staff evaluations. One participant suggested that the list be cross-referenced with current research on effective leadership to look for items to eliminate or to add.

Group 2 disagreed with the tool's statement that the evaluator should be responsible for gathering evidence. Rather, most participants agreed that the administrator should be responsible for providing evidence. One participant strongly disagreed with the exclusion of student attendance and discipline data. Participants agreed that the observations and artifacts listed were important, but agreed that other evidence could be added. Several participants noted that the focus should not be on provision of evidence, rather it should be on the principal's use of data; providing evidence does not mean that the principal is proficient in the use of the data, only in the accumulation of data.

Rubric

Neither group reviewing the Indiana RISE System was positive about the system. The two groups cited the tool's lack of adherence to Michigan standard rating language, and cited the length and vague language in the rubric as drawbacks. Both groups wondered why the evaluation was not linked to the local school improvement plan and noted a lack of flexibility in this regard. Neither group liked the letter grade assigned as a summative rating, although the second group was much more vocal about the letter grade than the first.

Participants in Group 1 felt lack of attention to management and communication skills in the RISE rubric was a significant deficit. The participants in the group felt that these areas were critical to the effective

functioning of a principal, should be defined and should carry more weight in the rubric. The Group 1 participants discussed implementing a second evaluation system to cover these critical areas, should the RISE rubric be adopted by the State.

Some participants in Group 1 noted that assigning a score would be difficult due to the imprecise terminology used in the rubric. However another participant liked some ambiguity: *I'm okay with that loose definition—looseness allows conversations to start and you can define what it means for your district.* Participants noted that subjective definitions could impact consistency in implementation across the district and State, and that consistency would become a problem when the ultimate ratings (the letter grades) are reported across the State. One participant stated that instead of the 'Student Learning Objectives' included in the rubric, the tool should be tied to the local school improvement plan: *Why wouldn't you use your school improvement plan? Why would you have competing goals?*

Participants in Group 2 were much more vocal in their criticism of the rubric, citing the rubric's lack of concrete definitions, and the problems with use of consistent language across the rating scales. Several participants felt that implementation would be difficult given the lack of specificity in the language: *Going back to the descriptors—they are too global, there is a lack of specificity, a lack of best practices; some of the content is ok, but many cases are not clear and are not scaled consistently.* One participant noted that the rubric is inconsistent with Michigan law. One participant stated that although not perfect, the rubric is *something to work with*, and went on to suggest that perhaps the domains should be defined based on the school improvement plan or district initiatives.

Across both groups, the majority of participants in the group felt that flexibility in application and use of the rubric would be needed to meet the needs of the variety of districts in Michigan. Participants in Group 1 felt that a self-assessment and goal-setting conversation should be added to the rubric. The length of the rubric was also cited as an issue in both focus groups. Participants in both groups used the term *checklist* to describe the rubric. In Group 2, one participant stated: *When I look at the spirit of the domains, I see a lot of compliance-based stuff.*

Participants in Group 1 did not extensively discuss the use of the letter grade as a summative assessment. At least one person expressed dissatisfaction with the letter grade, while another stated that she could understand that it was incorporated in order to arrive at a summative rating. One participant in Group 1 stated that he did not agree with the weights assigned to the elements of the rubric: *Weighting should be flexible by district—I don't think there is a one-size-fits-all for weighting it.*

The participants of Group 2 were strongly opposed to the use of a letter grade in the summative assessment. It was felt that the letter grade moved the tool out of the realm of a formative assessment toward a summative assessment of performance. One participant in Group 2 stated *I'm having a hard time going from the ambiguous language in the rubric, to the specifics on the scoring and grading. I have a hard time seeing this as a learning and growth tool.* Comments included: *The letter grade is a show-stopper; The letter grade is an effectiveness stopper.* At least two participants in Group 2 expressed concern that the letter grade would discourage emerging administrators. One participant in Group 2 noted that the school letter grade was heavily based on state assessment results, which unfairly targeted districts with high numbers of low socio-economic pupils, or districts with high numbers of English-language learners. Another Group 2 participant stated that Indiana was removing that aspect of the summative assessment from the tool, and questioned why Michigan would consider implementing the tool as shown.

Feasibility/Implementation/Training

Participants in both focus groups agreed that it would take a significant investment of time for users to develop an understanding of the language of the rubric: *There will be a lot of internal conversation between supervisors and principals to build a common understanding of what this looks like.* Concern was expressed that School Boards responsible for using the rubric would have difficulty with the terminology used in the rubric.

Participants in both groups agreed that flexibility and adaptability would be needed in order to make the rubric “fit” the variety of roles, building needs and districts across the state. One participant commented: *We could implement it as is, but it would be compliance.* Another participant stated: *We would have a mis-fire if we implemented this as is.*

Participants in both groups suggested that the ISDs/RESAs could play a role in implementation, by providing a centralized training resource and providing “train the trainer” workshops. In Group 2, the use of the letter grade as a summative assessment considerably raised the bar for the expectation of the amount of training that would be required. One participant described the need for training to *lower the threat level for principals.*

Participants in both groups stated that the State should provide funding for training activities: *Don’t create another unfunded mandate; We don’t have the resources for this.* Participants in both groups stressed their desire for a meaningful evaluation process that would encourage growth and development: *Don’t make this something that we check off, we want it to be a meaningful process; We don’t shy away from accountability—but we want a model that supports growth, that is valid and reliable.*

Other

Participants in Group 1 felt that the Michigan Standards for the Preparation of Principals were outdated, and should be revised.

Several participants in Group 2 requested that the State provide more than one evaluation tool from which to choose. If providing more than one tool is not possible, Group 2 participants asked for flexibility and adaptability in the tool selected.

Marzano School Leadership Evaluation Model

Domains

Group 1 participants were for the most part comfortable with the domains and subdomains in the Marzano Model, stating that the domains provide a clear picture of expectation for an instructional leader. One participant described the domains as providing a *360 degree picture* of instructional leadership. *This nails instruction* commented another participant. Another described it as *thorough but intense*, commenting both on the scope and the length of the domains and subdomains. One participant described the domains as *manageable*. Another Group 1 participant described the domains and subdomains as *what we all aspire to* as instructional leaders. Participants in both groups commented on the tool's strong connection to research.

One participant in Group 2 noted that the Marzano MModel was less overwhelming than some other tools, and was supported by research; *It fits on one page, it is not so overwhelming that you lose focus*. The participant noted that the tool could be used to focus the principals on instructional leadership. One Group 2 participant liked the change in focus (away from building management), however another Group 2 participant noted that principals will need support for building management if their focus was to be solely on instructional leadership.

There was strong agreement in both focus groups that the Marzano tool lacked domains reflective of a principal's critical and daily responsibilities for building management, which take up a large part of the day. *This is a wonderful tool for measuring the effectiveness of an instructional leader, but there is a lot more to being a building principal than being an instructional leader* stated one Group 1 participant. The broad consensus in both groups was that the Marzano Model was incomplete without incorporation of these critical building management domains, including planning, budgeting and fiscal management skills.

Participants in both groups also felt that the Marzano tool lacked a focus on communication skills which both groups felt were critical for a building manager. Participants commented on the principal's need to resolve conflicts, hold critical conversations and develop rapport with staff, parents and students. A Group 2 participant described the 'Climate' domain as *fluffy and vague*, and felt that it did not cover the communication skills adequately.

Some Group 1 participants commented on the length of the list of the domains and subdomains, describing them as *overwhelming*. The participant worried that the principals would be trying to "hit" everything in the list of domains and subdomains. While another Group 1 commenter felt that the contents of the domains were good, the tool was *not realistic*. The participant suggested that evaluators and principals focus on one or two domains in a year. Narrowing the focus would allow more attention to be given to the domains of focus. One participant in Group 1 suggested adding yearly goals and a self-assessment.

Participants in both groups noted that some domains would need to be modified for the various roles and responsibilities across the districts. A member of Group 2 commented that the tool did not seem to emphasize the principal's responsibility for teacher evaluations. One participant in Group 2 felt that teacher evaluations should be covered in a separate domain. One Group 2 participant felt that the domain covering 'Curriculum Development' would need to be removed or revised; principals implement curriculum but do not develop it.

Concern was expressed in Group 2 about how the domains would be weighted. Discussion in the group favored having the districts control weighting. Districts may need to change weighting from year to year to accommodate local conditions or priorities.

Indicators and Behaviors

Participants in Group 1 felt that indicators needed more focus on building management and developing and maintaining relationships with people. One participant commented that effective administrators have systems, processes and procedures in place to help with management, in order for the focus to be on

improving student achievement outcomes. However there was consensus in the group that those activities that lead to effective building management were not reflected in the indicators and behaviors. The group felt that the indicators did not reflect on, or evaluate, areas such as management and communication where administrators often struggle. *If you are not a good manager, you will not be successful in creating a learning environment.*

There was no agreement in Group 1 on the clarity of the tool. At least one participant commented on vague language in the indicators, noting that the meanings of terms must be made *crystal clear*. While one participant commented that the thresholds for performance are good, another participant noted that the thresholds for moving from one level to another are not totally clear. The participant felt that the Marzano Model is not as clear as the Danielson System. Participants in Group 1 noted that the rating language in Marzano differed from that required by the State of Michigan, and that some guidance must be given in order to translate the Marzano rating language to the State of Michigan rating language.

Participants in Group 2 felt that the indicators were not concrete enough to be helpful, or to show a principal how to move from one level to the next. Verbs and adjectives were felt to be vague, and definitions were needed for words like “regularly”, “rigorous”, “critical”, etc. One participant asked for descriptions that would let the principal (and the evaluator) know what good performance *looked like, sounded like, felt like*. The participant went further to state *We need a system that will help develop principals and show them concretely how to move from one level to another --this is not concrete enough in terms of descriptors*. At least three participants agreed that the Danielson tool was much better at providing clear, concrete guidance.

Evidence

Group 1 described the list of possible evidence as *comprehensive*. A large part of the discussion focused on the need for flexibility to choose evidence material to match the administrator role, building or district situation. One Group 1 participant noted that Dr. Marzano was very vocal about not liking adaptation of his tool, which raised a serious concern about the usefulness of the Marzano evaluation tool. One Group 1 participant commented on the incorporation of student growth and achievement in the tool, noting that the method of determining student growth in the State of Michigan has not been decided. The participant felt that it was hard to see how the student growth determination will connect to the domains listed in the rubric.

Group 1’s discussion of evidence reiterated that the domains do not include management and communication skills. Group 1 participants agreed that the Marzano tool as it was shown would not allow for evaluation of those critical skills.

Group 2 described the list of possible evidence as *overwhelming*. Several participants agreed that the list of evidence would become a *to-do* list for principals, leading principals to spend valuable time attempting to check all of the items off the list. Group 2 participants asked for information about which artifacts were the most important (supported by research), and if artifacts would be weighted. At least three participants indicated that the guidance on evidence in the Danielson tool was better than that found in the Marzano tool.

Rubric

Participants in both focus groups felt that flexibility would be needed to adapt the tools for various roles and responsibilities, building and district configurations found in Michigan.

The consensus in Group 1 was that while the Marzano tool is *comprehensive* for instructional leadership, it did not provide a means to evaluate the critical domains of management and communication. While one person thought the indicators were clear, at least two other participants thought more clarity in the language of the rubric would be needed. One Group 1 participant described the rating labels (the levels of performance) as confusing. Participants felt that some translation would be needed to move between the rubric and the State of Michigan rating language. Participants in the group agreed that the research behind the tool and the focus on instructional leadership were the best aspects of the tool.

Participants in Group 2 did not like the tool, describing the tool as *seriously lacking in concrete definitions* and *too wordy*. The participants in Group 2 felt that the rubric was too long, and would require too much time from evaluators and principals, both of whom would attempt to cover all of the items listed in the rubric. Participants asked if some domains were more important than others, and whether or not they could choose some of the domains. One participant wondered if the Marzano 9 or the Marzano 3 could be emphasized in order to reduce the length of the rubric.

One participant in Group 2 noted that it was difficult to differentiate between Innovating and Implementing; in some cases the lower-level descriptors were more rigorous than those for the higher level. Another participant in Group 2 wondered if the Marzano performance rating labels would be more difficult to use than the State of Michigan rating labels. There was agreement in Group 2 that the tool was long and would be difficult to use to demonstrate growth over time. *The way this is presented is overwhelming and doesn't show how to progress to be the best*, stated one participant.

Concern was expressed in both groups regarding the method of calculating student growth and achievement, and how student growth and achievement calculations would fit into any tool selected by the State.

Feasibility/Implementation/Training

Adaptation for different roles and responsibilities, and for building and district goals, was discussed in both groups. There was common agreement in both groups that one tool would not fit all districts and flexibility to adapt for local needs and priorities would be necessary. A management domain would be a critical adaptation for Group 1, as would adaptations of rubric items and evidence for the various roles within the districts. The tool would need to be adapted for consistency with the State of Michigan rating language.

The length of the tool raised feasibility questions in both focus groups, with Group 2 expressing more concern than Group 1. Both groups requested input or a workshop from a current user of the system to provide a hands-on demonstration of the use of the tool.

Both groups agreed that the feasibility of the Marzano tool was difficult to determine without information on the cost of the system. Funding is an impediment to implementation in both groups. Some districts are currently in deficit, or projecting deficits, and would not have funding for procurement, training, or implementation of the tool. *Feasibility depends on costs; We need to have adequate funding to implement this correctly* were comments from one participants. One participant in Group 1 suggested that communities would push back if districts were required to implement a system without financial support from the State; others in the group agreed with that statement.

Both groups agreed that technology would assist with implementation, but that funding for technology is not currently available. At least one person in each group felt that an electronic application was necessary in order for the system to be feasible. Both groups worried about ongoing licensing costs for use of the tool and how increases in licensing cost might impact district budgets over time. One participant worried that a paper implementation of the tool might not achieve the fidelity required by Dr. Marzano.

Both groups agreed that training in the use of the tool would need to be intensive, spanning multiple days. Inter-rater reliability could not be achieved without in-depth training. Comments included: *A critical mistake on the State's end would be to tell us in July that we have to roll this out in August--that will never work; Take the time to do it correctly*. Each group suggested that the ISDs/RESAs could help coordinate and provide training for the system. Both groups felt that a planned roll-out, with sufficient time for training and understanding of the system would be needed.

MASA School ADvance Administrator Evaluation System

Domains

Participants in both focus groups commented on the ordering and presentation of the domains in the School ADvance tool. Participants in both groups shared concern that the ordering and presentation suggested importance and weighting of the domains. One participant in Group 1 questioned whether the rubric placed too much emphasis on student results, however another participant commented *emphasis on results is the reality that we have to deal with*. Group 2 participants also felt that the ordering of the domains, with Results at the top, sent a strong message that the end result was more important than the means to achieve the result. One Group 2 participant stated: *What is put first on the list sends a message*. Another: *The focus is on the ends, not the means*. One Group 2 participant suggested rather than having a domain titled 'Results', that results should be embedded in each of the other domain areas. Another Group 2 participant wanted to visualize the domains as a flow chart, with the leadership, processes and systems that lead to results. Participants in Group 2 worried that having results at the top of the list would not lead to conversations on leadership and management skills necessary to achieve growth. A participant commented *I would focus on leadership first, and move results to the bottom*. There was consensus in Group 2 that the ordering and presentation of the domains should be changed.

There was broad agreement in Group 1 that achievement of good student outcomes is dependent upon good management skills, and that management skills are not reflected sufficiently in the domains of the rubric. Participants in Group 1 agreed that the domains should be rearranged, or weighted more heavily in favor of effective and efficient operations, systems and processes, and school climate and communication skills, all of which were felt to ultimately lead to better student outcomes. Many districts represented in the Group 1 focus group have principals that fill other roles in addition to building management, such as transportation, cafeteria management, athletics, and so on. The participants agreed that the depth and breadth of the other tasks completed by principals was not reflected in the domains and subdomains. One participant commented: *The scope of what a principal deals with is not in here*.

Participants in Group 2 also felt that management tasks were not well-covered in the domains and subdomains, but additionally mentioned other areas they felt needed more emphasis. One participant in Group 2 felt that the principal's responsibility for conducting teacher evaluations was not emphasized enough in School ADvance. That participant felt strongly that teacher evaluations, and knowledge of appropriate supervisory practices and legal implications was not addressed sufficiently in the domains and subdomains.

Some participants in both groups felt that communication skills and building relationships with staff should be emphasized more in the domains and subdomains. One participant in Group 1 mentioned that building staff morale was important, and that staff morale impacted all other domains. In Group 2, there was discussion of embedding these items in each domain, or perhaps adding a domain for these types of skills.

At least two people in Group 1 commented on the length of the list of the School ADvance domains and rubric, expressing the desire to reduce the number of blocks by ranking the most useful domains. One participant described the list of domains as *overwhelming*. In contrast, Group 2 participants stated that they appreciated the brevity of the list of domains and subdomains. One participant called the list of domains *manageable*.

Indicators and Behaviors

There was no consistent theme that emerged from Group 1 about indicators and behaviors. One participant suggested that the rubric was "harsh" and that the expected behaviors were too rigorous given that most principals fulfill more than one job in the district. One participant noted that not all indicators were appropriate for their districts, and evaluators would have to modify those cells to fit the district. The group agreed that it would be difficult for an evaluator to observe behaviors in all of the cells in the rubric.

One participant noted a lack of indicators for the 'Results' domain. While one participant felt the indicators and School ADvance tool was easy to use, another felt it was too long. One participant felt that the rubric provided clear targets; another noted that the indicators were more descriptive and allowed some subjectivity on the part of the evaluator. The group agreed that School ADvance System was far less subjective than purely narrative evaluations.

Members of Group 2 were more positive about the tool. One person commented *those things that we know make the difference in leadership are embedded a little in each of these areas*. At least two people stated that the indicators provide enough guidance to help a principal understand what is needed to move from one level to another. One participant thought it was difficult to make some judgments, but that this system was better than others. Some participants in Group 2 commented on what were perceived as subjective descriptors, stating that the districts would need to spend time defining those terms in order to achieve consistent application. Another participant felt that the subjectivity was good, and allowed districts the flexibility to incorporate district-level expectations for performance. The members of Group 2 agreed that the most difficult area to discern in any rubric is between the "Highly Effective" and "Effective" rating. One Group 2 participant wondered whether the School ADvance System lined up with the accreditation standards, and whether leader responsibilities for accreditation were included in the indicators. Another Group 2 participant liked that the Results domain allowed the flexibility for building-level goals, and for establishment of Board-derived goals.

Evidence

All participants in Group 1 liked the inclusion of a self-assessment as evidence. There was some discussion about the indicators of student growth and achievement. While the tool allows for different tests to be used as evidence, participants noted that their districts did not have funding to pay for other tests. Most participants welcomed the evidence suggestions and felt that the evidence would be helpful in terms of documentation. *This is a step in the right direction* commented a Group 1 participant. Another participant wondered if the focus of the evidence was in the right place, noting the emphasis on student achievement results. Another Group 1 participant would have liked more evidence of management routines, processes and procedures.

Group 2 participants were more positive about the suggestions for evidence. Most members agreed that the self-assessment was an important piece of evidence. Participants using technology to manage documents and evidence felt that it was very helpful. Participants suggested that other items could be added to the list, which would allow a district-based focus.

Group 2 participants agreed that the focus should not be on the evidence itself, but on how the principal uses the evidence or data to move the building to the short and long-term goals. *We don't want this all to be about gathering the evidence, we need to have that discussion about growth* stated one Group 2 participant, with agreement from the members of the group.

Rubric

Participants in Group 1 felt that the inclusion of a rubric in an evaluation system was a good step, and would help to make the process less subjective. One participant described the School ADvance rubric as complete and easy to use and another stated: *It is good that there are clear targets*. However, another participant noted: *This is not a magic bullet that will make our students do better*. One participant described the rubric as overwhelming and another stated that he would like to see the system streamlined. The Group 1 participants agreed that management skills should be added to the rubric, with one participant noting *it needs to capture the full range of a principal's performance*.

Participants in Group 2 were much more positive about the rubric. Statements included: *It flows well; the rubric is pretty easy to follow in terms of setting the expectation and showing how to meet the standard; more straightforward than what we are using now; it's not easy to judge, but this tool is better than others*. Participants agreed that achieving consistency across evaluations could be an issue due to some of the undefined descriptors used in the rubric. It was stated that an evaluator would need to be careful to apply the rubric consistently across principals, and consistent application across districts would be difficult to achieve.

Group 1 participants briefly discussed the summative rating and weighting of the domains. Several participants in Group 1 questioned the weighting of the domains. More than one participant suggested that the districts should choose the weights of the domains based on local goals. One Group 1 participant felt that the directions for creating a summative rating needed to be clarified. On the other hand, a participant in Group 2 liked the clear guidance on scoring to arrive at a summative rating. In Group 2, there was no consensus on the suggested weighting of domains or elements. However all members of the group endorsed having the emphasis on the conversation about growth, and not on the score at the end. *We want a tool that is formative rather than summative* commented one person in Group 2, with endorsement by all participants around the table.

There was agreement in both groups that districts would need to have flexibility to add locally-established goals. In Group 1 there was discussion of adaptation of the rubric for local conditions in each district; some districts would not have any activities in the highly effective category due to lack of district resources. At least two participants in Group 1 noted that the school board members with responsibility for completing evaluations will not be able to complete the rubric without training.

There was discussion in Group 1 about working with principals to help them understand the ratings of “Highly Effective” and “Effective”. Superintendents will need to work with principals to create the expectation that an “Effective” rating is acceptable and normal, and that very few individuals would be expected to receive a “Highly Effective” rating.

Feasibility/Implementation/Training

Participants in both groups felt that implementation of the School ADvance tool would be feasible, although Group 2 participants felt that larger districts with more principals would have a more difficult time due to the length of the rubric. One participant in Group 2 noted that the tool works with STAGES, which is already in use in the district. Group 2 discussed the use of electronic document management systems. Such systems were felt to increase efficiency and be helpful for larger districts with many administrators, but not necessary for smaller districts. While many participants in Group 2 agreed the system was not perfect, they felt that it could be adapted for use with modifications that included adding district-level goals or a narrative component. One participant in Group 2 felt that since the tool looked and felt like the Danielson System, it would be more readily understood and accepted. One participant in Group 1 felt the tool would not be hard to implement.

Some participants in both Group 1 and Group 2 worried about the length of the rubric and the amount of time it would take a superintendent to complete evaluations. One participant in Group 1 suggested adding “focus areas” during the evaluation, and another suggesting adding a narrative box similar to that used in the Danielson tool. Another Group 1 participant felt that the tool needed to be streamlined.

Participants in both groups agreed that implementation would take time. Users in the districts would need to discuss the rubric and come to a shared understanding of the language. Participants in both groups agreed that that piloting the tool would be helpful. One participant in Group 1 liked the phased roll-out plan mentioned in the documentation for the tool. Several participants in Group 2 liked the idea of having the ISDs/RESAs coordinate training. It was felt that the ISDs/RESAs could provide ongoing expertise and user support for the tool. Group 2 participants felt that centralized training would assist with establishing inter-rater reliability and consistency in application within and across districts. One Group 2 participant suggested that it would be helpful to hold a day-long training with evaluators and principals working together to develop a shared understanding of the tool. Another Group 2 participant suggested one-on-one meetings between the evaluator and the principal.

Funding for implementation and training was mentioned multiple times in both focus groups. Participants in both focus groups agreed that the State should provide funding for training and implementation. Funding for the tool and for training could be a significant barrier to implementation in some districts. One Group 2 participant stated that *all evaluation requirements coming from the State deserve more funding to get the job done*. At least two participants in Group 2 stated that expense for licensing, training and implementation should be shared with the State.

Participants in both groups felt that the School ADvance tool would need to be adapted in order to evaluate the various roles found across the districts. Participants in both groups felt that flexibility and adaptability to incorporate building and district goals would also be needed.

Reeves Leadership Performance Matrix

Domains

There were some positive comments about the domains and subdomains in both groups. However, participants in the second group were more positive about the tool than the first group. Participants in both groups made general comments about liking the layout and the completeness or *broad scope* of the domains: *I like the way this is laid out; Great deal of completeness and is clearly related to improving student achievement.* One person commented that this tool was better suited for principals than the Marzano or ADvance tools, due to its limited focus on curriculum, which is often handled at the district level. Both groups noted that the domains focus on student achievement, and both groups commented that there was no guidance from the State on how to measure achievement. Both groups noted that the domains would need to be modified for use across administrator positions and districts.

Group 1 participants found the list of domains far too long, and felt that it would take an evaluator far too much time to complete an evaluation: *This will take an exorbitant amount of time to do thoughtfully; How can I say that an organization has grown over time with so many factors to look at?* At least one person felt that the list could be used, and the descriptions were flexible enough to cover most areas of a principal's responsibility. Other participants in the group focused on the length of the tool. There was agreement in the group that the tool did not focus enough attention on interpersonal or "relationship" skills, such as communication, collaboration, and school climate. At least one person mentioned that the list of domains should tie in to domains covered in teacher evaluations in order to provide vertical aggregation and provide a view of the *whole*. Participants in Group 1 mentioned communication skills, teacher evaluations, management, and budgeting as items that were missing in the list of domains. There was general agreement in Group 1 that student achievement and communication were the most important domains. There was no agreement on the relative importance of other domains. While one person in Group 1 specifically mentioned liking the addition of technology as a category, another in Group 2 commented that technology should not be included given that not all schools had funding for or access to technology: *You can't evaluate someone on something they don't have. I don't know how this works.*

Group 2 participants were more positive about the domains. Unlike the first group, at least one participant felt that there was strong emphasis on interpersonal skills. This group identified the tool as moving *180 degrees away from the management function*. However, Group 2 participants felt that domains associated with management were critical parts of a principal's performance, and must be added to the tool. While one participant felt that some of those key management items could be covered under 'Time, Task and Project Management,' others agreed that the task of management was not well reflected in the list of sub-domains. *Lack of attention to the management function leads to an administrator losing his job very quickly* was one comment that obtained widespread agreement in Group 2.

Members of the second group also commented on the length of the list of domains, but felt that the length could be addressed by setting priorities by person or building: *I love the traits listed here, but when you look at applying a system like this, you'd probably extract items into a specific plan; I would customize this by need.* There was no agreement on the level of importance of certain domains. There was some discussion as to whether or not the domains reflected the growing importance of the amount of time a principal spends on teacher evaluations. One Group 2 participant did not feel that the teacher evaluation task was important to include, while another felt that the responsibility for teacher evaluations should be called out specifically in the domains.

Both groups questioned the methodology for weighting the domains and subdomains. There was a broad discussion in the second group regarding weighting of the domains, and how that weighting would impact the superintendents' ability to remove a principal who lacked critical management or communication skills or demonstrated a lack of professional integrity. Group 2 participants felt that skills or behaviors (such as those in the domains of 'Personal Behavior', 'Communication', and the missing management domain) were felt to be so essential that they might require heavier weights than other domains; *How do you weigh the Achilles heel?; Failure in one domain can have such an adverse impact to the organization that it will cause job loss--how do you do that with this rubric?* All agreed that those skills and behaviors must

be included in the rubric if they were to be evaluated, and if they could potentially lead to the termination of a principal.

Indicators and Behaviors

Concern was expressed in both groups regarding the number of indicators and the amount of time it would take to complete an evaluation. Participants in Group 1 liked the mention of alignment to the school plan and student achievement, and linking to the school vision. A Group 1 participant commented: *I think we can use these subdomains to guide us to increasing our capacity.*

Participants in Group 2 felt the indicators were clear: *This is so clear, it makes it easy.* However there was some agreement in Group 1 that indicators for some domains and subdomains were inconsistent. One Group 1 participant felt that the cells did not build upon each other to show consistent growth from the lowest to the highest rating. Others in the group agreed. One Group 1 participant noted that some Effective indicators appeared to be more stringent than those for a “Highly Effective” rating under the same subdomain. Participants in both groups agreed that the distinctions between the “Effective” and “Highly Effective” categories were the most difficult to judge.

Both groups commented that the rating language used in the Reeves tool is inconsistent with the language now required in the State of Michigan. Concerns were expressed in both groups regarding the validity of the tool if the rating language were to be changed.

Evidence

Both groups explored the different forms provided in the Reeves tool and discussed how those forms could be used to solicit or provide evidence to document performance. Participants in both groups mentioned that the coaching form provided some flexibility in terms of use and in terms of defining what evidence might be brought forward to document performance. One participant suggested that the templates provided might be used as a self-assessment. Another participant asked for specific examples of evidence that were best-practice and research-based.

Both groups mentioned that some of the suggested evidence would be difficult to observe, and mentioned that considerable time could be expended (either on the part of the principal or the evaluator) to collect and document evidence. Participants in both groups suggested having the principal define or submit evidence as needed to document performance (rather than having the superintendent have that responsibility). Participants in both groups mentioned customizing the evidence to a building or person-level need. One participant suggested that it would be helpful to establish an annual timeline for carrying out the evaluation/observation tasks.

Rubric

Many participants in both groups felt the rubric was clear and easy to understand, and provided clear guidance to participants on how to move from one level of performance to the next: *Yes, definitely clear; This is easily understood--you could make this into a job description; this is easier to make a judgment because it is objective.* However at least one person in Group 2 felt that the language in the rubric was open to interpretation: *Everyone who reads it will understand it, but the danger is that they will have many different interpretations.* Another participant felt that the ambiguity in the rubric language provided more leverage for the evaluator than did the Danielson model, due to the lack of specific descriptors such as “never” and “always.” The first group questioned how the rubric could be adapted for years of experience: *Would the rubric work as well for a new administrator as for one with many years of experience?*

The length of the rubric was an issue for many participants in both focus groups. Participants in both groups commented that it would take a significant amount of time to review each domain and subdomain with employees: *This is pretty cumbersome; It would be really difficult to go through every single one of those domains and subdomains.* Participants in both groups talked of adapting the tool for districts and buildings. Both groups discussed customizing the tool by focusing on three or four domains in the rubric for a specific district or principal, and thereby reducing the amount of time required for the evaluation. One person suggested that the short-form might be the best form to use for conducting an evaluation.

However, participants in the groups also wondered about the validity and integrity of the tool after customization.

Feasibility/Implementation/Training

Feasibility and implementation issues were discussed throughout both sessions, regardless of the prompt presented. Discussion largely revolved around managing the scope and length of the rubric, and time involved in gathering evidence. Both groups discussed adapting the tool to the needs of the role, principal or district. Some domains would be dropped for assistant principals who focus on one area such as discipline, or other administrators such as the special education director or athletic director.

The lack of implementation instruction for the tool was noted by participants in both focus groups. Participants in the groups overwhelmingly indicated that *a great deal* of training would be needed both for the evaluators and the principals. Users will need to understand the terminology involved, the evidence required, and the means to arrive at a summative rating. One participant suggested that a pre-organized structure was needed for training so that everyone would come to the same understanding of the ten domains included in the tool. Participants in both groups suggested that training needs would be intensive, and that the tool could not be rolled out quickly. Both groups felt that the ISDs/RESAs could be very helpful in rolling out a new tool, both by providing training opportunities, train-the-trainer sessions, and working with the districts to obtain inter-rater reliability. One participant noted that inter-rater reliability would be difficult to achieve without training.

Both groups felt that implementation of the tool could be customized through the use of self-evaluation, dialog, and alignment with district goals, ultimately reducing the overall focus to a few domains: *Instead of the shotgun approach, this could be customized to a building or to a need. If you have to go through all of these [domains], the effect gets watered down and it might not move the building forward. I would recommend you focus on specific areas not all domains.*

It was suggested in one group that implementation could be done by focusing on a few domains each year until all domains had been covered, thereby emphasizing growth over time. Participants in both groups mentioned piloting, and phased-in approaches to implementation. One participant suggested using the current district tool in parallel with the new tool for one year in order to allow the principals to become accustomed to the rubric and the language. Participants in both groups overwhelmingly indicated that they could work with any tool selected by the State. However participants in both groups emphasized that they would prefer a tool that emphasizes growth and development, as opposed to documentation.

Participants in both groups indicated that funding for the tool itself, for technology to implement the tool, and for training would be needed. One participant noted that use of technology such as STAGES or Teachscape would make implementation and use much easier, however funding would need to be provided for this technology: *We don't have the resources to implement this and do a good job.*

Other

At least one participant in each group noted that, in their districts, salary increases for principals are not tied to performance. These participants felt that the salary-increase system for principals should change to a performance-based system.

Participants in one group expressed a desire for “exemptions” from the use of the State-selected tool, if a locally-used evaluation tool meets the requirements of the State. A participant in the second group stated that it would be helpful if the districts could choose from more than one tool. The availability of several models would allow the districts to choose a model that is more applicable or useful to local conditions.

One participant suggested that the MCEE consider the Thoughtful Leader tool. This aligns with the Thoughtful Teacher tool that is being piloted in some districts. One participant noted that the Reeves Matrix is *stronger and more focused than Marzano*, and that *School ADvance has some nice features, but is more complex than Reeves*.

Appendix H: Validity and Reliability Required for Waiver Approval

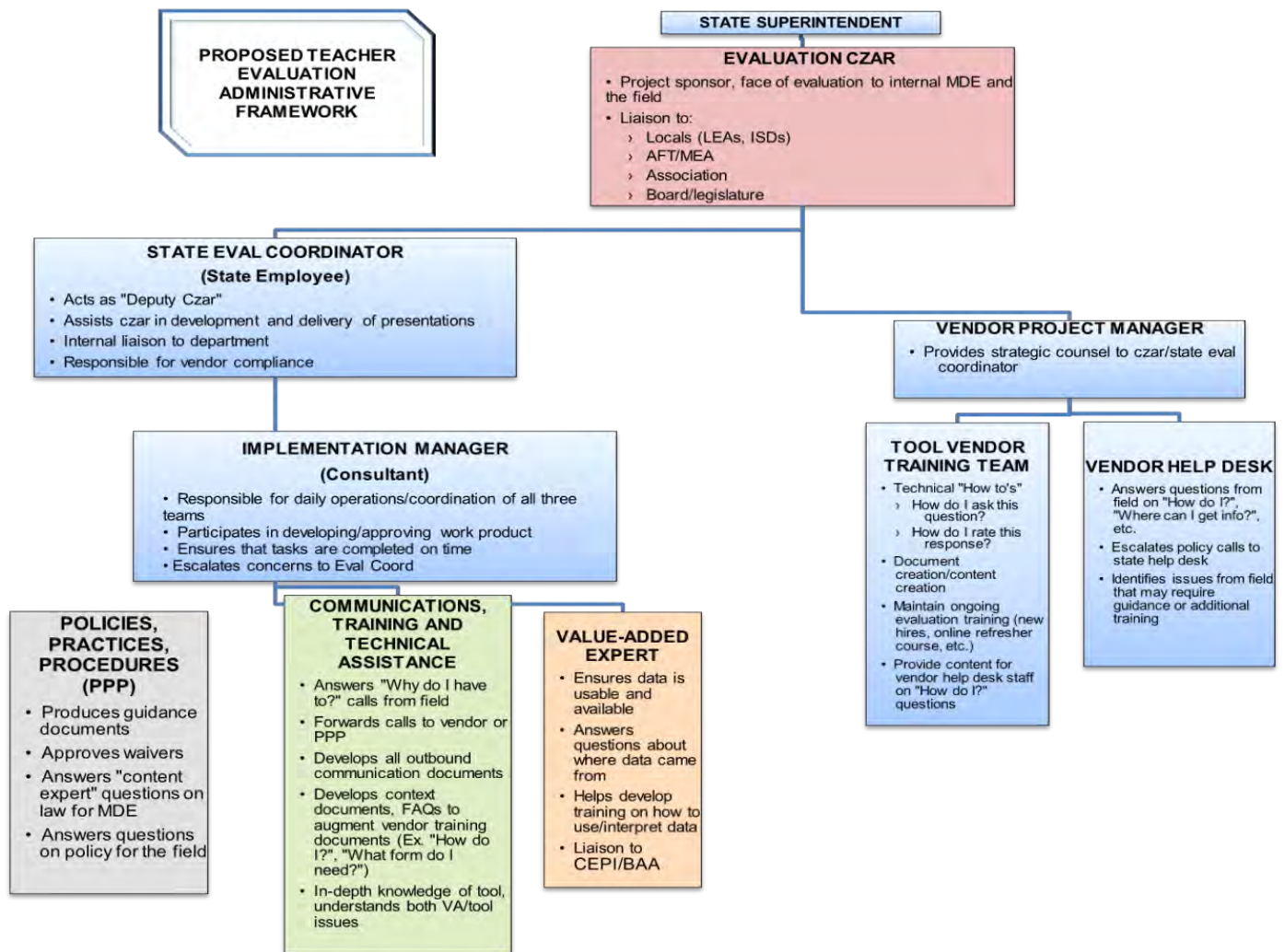
Construct (content) validity: The items on the observation instrument should cover all of the domains on one of the instruments piloted by MCEE. This can be demonstrated through mapping how items on the locally developed instrument correspond to items on one of the MCEE piloted instruments.

Convergent validity: The scores a teacher receives from observations on the locally developed instrument should be positively correlated to scores the same teacher receives on one of the MCEE piloted instruments. This can be demonstrated through local research or other published research, and will consist of a correlation matrix having a sufficient number of cases ($n > 50$).

Predictive validity: The scores teachers receive on the locally developed instrument should be positively correlated to measures of student growth in learning. This too can be demonstrated through local research or other published research, and will consist of a correlation matrix having a sufficient number of cases ($n > 50$).

Reliability: A locally developed instrument will be judged to be reliable only when both of the following conditions are met: (a) the measure of teaching quality derived from the observation instrument has a reasonable internal consistency; and (b) there is evidence that when two observers use the same instrument at the same time to observe a lesson, they will arrive at roughly the same scores for that lesson. Evidence of (a) shall be provided through standard measures of internal consistency (e.g., Cronbach's alpha, factor analysis, etc.). Evidence of (b) shall be presented through standard measures of inter-rater reliability such as kappa or another accepted measure of agreement among raters. In addition, those proposing to use a locally developed measure should present evidence on the precision of their measures by describing how a standard error of measurement will be calculated and showing how that standard error takes into account and is a function of the number of raters and occasions of measurement. Any evidence of reliability should be based on a sufficient number of cases ($n > 50$).

Appendix I: Possible Infrastructure Frameworks



CREDIT: Emily Houk, Public Sector Consultants

**Proposed Organizational Chart
Submitted by the Michigan Department of Education**

In order to adequately carry out the responsibilities outlined in this report, significant new appropriations will be needed in the following areas:

- New contracts with external entities
- New staff to oversee those contracts
- New staff to oversee modified and new state data systems
- New staff to project manage all aspects of educator evaluation systems to achieve a comprehensive and coherent system delivered on time both in the development phase and each year in the operational phase

While these new resources are necessary to adequately carry out the activities outlined in this report, it is also possible to leverage existing staff to reduce the number of new staff needed. The list of contracted services necessary is given below:

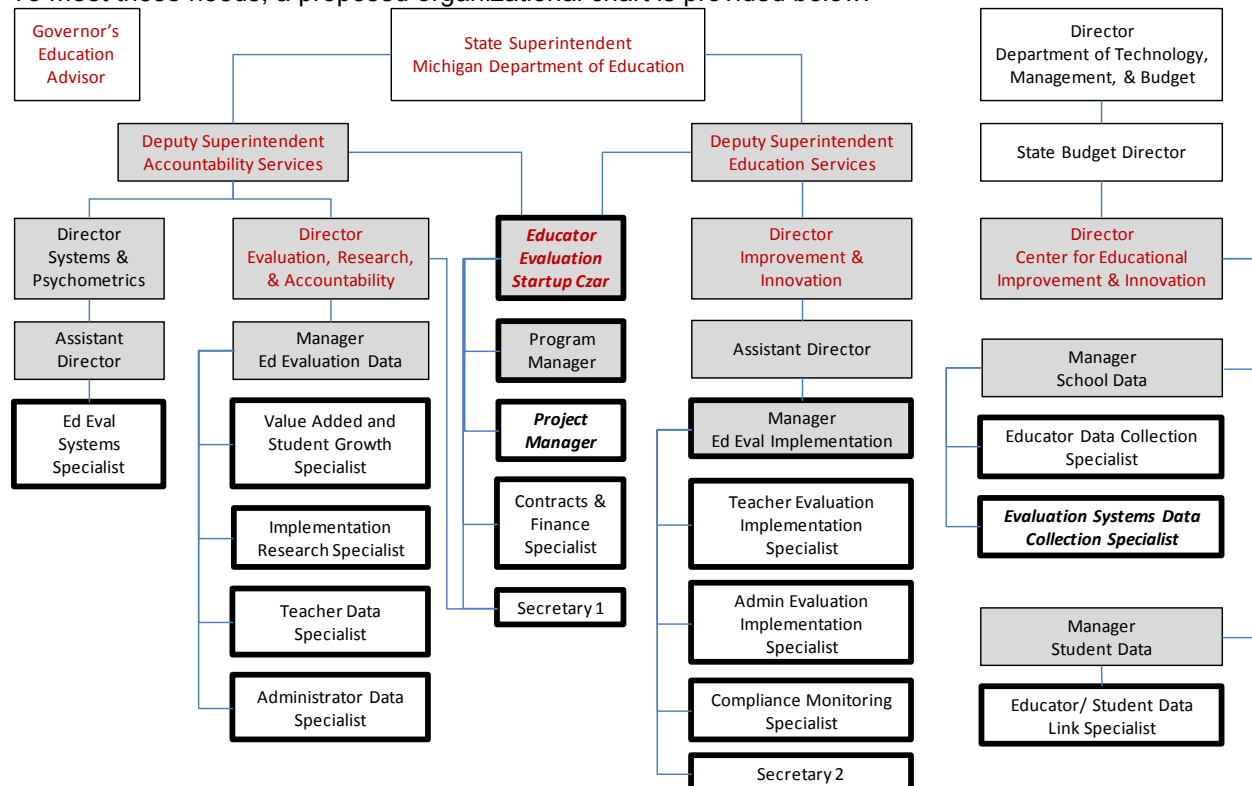
- Student growth data provision (already in existing contracts)
- Provision of new assessments in non-tested grades and subjects (already in existing contracts, but would require new appropriations for administration)
- Training on the use of student growth data
- Value-added modeling services
- Teacher observation system services
- Administrator observation system services
- Compliance monitoring services
- Evaluation research services
- State data system builds and enhancements services

In order to adequately oversee the performance of these contractors and sustain the new and modified systems, the following new staff roles are needed:

- Rapid development of RFPs and contracts, with ongoing contractual and financial monitoring
- Management of a value-added modeling services contractor
- Management of the student growth data portion of assessment contractor
- Management of a teacher evaluation contractor, to carry out the following activities:
 - Training on SLOs, teacher observations, using teacher observation data in evaluations, using growth and/or VAM data in evaluations, using other factors in evaluations, combining factors for overall evaluations, and using evaluation results to inform professional learning
 - Provision of the observation tool
 - Building/hosting state teacher observation data systems
 - Provision of teacher observation data analysis and reporting systems
- Management of an administrator evaluation contractor, to carry out the following activities:
 - Training on professional practice administrator ratings, using administrator professional practice rating data in evaluations, using growth and/or VAM data in evaluations, using other factors in evaluations, combining factors for overall evaluations, and using evaluation results to inform professional learning
 - Provision of the professional practice rating tool
 - Building/hosting state administrator professional practice rating data systems
 - Provision of administrator professional practice rating data analysis and reporting systems

- Management of a compliance monitoring contractor to do small-scale random evaluation of local system compliance with state law, and to annually report on compliance
- Management of an implementation evaluation research contractor to report annually on the quality, rigor, and consequences of local implementation
- Management of a state data systems enhancement contractor
- Management of new/enhanced state data systems
- Management of new staff
- Startup leadership
- Program/project management

To meet these needs, a proposed organizational chart is provided below:



In this proposed organizational chart, the following conventions are used:

- Bolded box outlines indicate new staff, where bold and italic lettering indicates temporary staffing needs (only during development years and the first operational year)
- Non-bolded box outlines indicate existing staff whose time would be significantly repurposed to manage educator evaluation system development and implementation
- Gray box fill indicates staff on a mid-level cross-agency management team to meet frequently to monitor progress on all development activities and to address issues as they arise
- Red lettering indicates staff on a high-level cross-agency executive leadership team to provide periodic executive review and executive leadership regarding development and implementation of educator evaluation systems

The total new staff needs are explained below:

- Ed Eval Systems Specialist – responsible for integration of all technical, data, and reporting systems relevant to educator evaluation across all relevant units

- Value-added and student growth specialist – responsible for directing student growth data provision by assessment contractors and directing the work of the value added services contractor
- Implementation research specialist – responsible for directing the work of the implementation research contractor, including many on-site visits for research purposes
- Teacher data specialist – responsible for directing the data analysis work of the teacher evaluation contractor
- Administrator data specialist – responsible for directing the data analysis work of the administrator evaluation contractor
- Educator evaluation startup czar – *temporary* staff responsible for statewide communication and outreach, policy development, and management of contractual, financial, and project management staff during development activities and the first year of implementation
- Program manager – responsible for coordinating all development and implementation activities across all contractors relevant to educator evaluation
- Project manager – *temporary* staff responsible for maintaining schedules and timelines during development and the first year of implementation
- Contracts & finance specialist – responsible for rapid development of requests for proposals, coordinating bid evaluations, and monitoring ongoing contractual, budget, and financial issues.
- Secretary 1 – responsible for providing clerical support to the startup czar's unit and the ed evaluation data unit staff during development activities and the first year of implementation
- Manager Ed Eval Implementation – responsible for managing all staff responsible for training and compliance monitoring
- Teacher evaluation implementation specialist – responsible for directing all training and technical assistance activities of the teacher evaluation contractor, including attendance at large numbers of training events
- Administrator evaluation implementation specialist – responsible for directing all training and technical assistance activities of the administrator evaluation contractor, including attendance at large numbers of training events
- Compliance monitoring specialist – responsible for directing all work of the compliance monitoring contractor, including attendance at large numbers of monitoring events
- Secretary 2 – provide clerical support for significant travel and documentation needs of the ed eval implementation unit
- Educator data collection specialist – responsible for directing the state-systems development work of the teacher evaluation and principal evaluation contractors and managing those systems thereafter
- Evaluation systems data collection specialist – *temporary* staff responsible for supporting the work of the educator data collection specialist and directing contractors enhancing existing systems to collect additional data regarding local district evaluation systems during development activities and the first implementation year
- Educator/student data link specialist - responsible for directing the work of systems modification/development contractors and managing the resulting systems thereafter (to include find-grained attendance capture, capturing proportional educator responsibility by content area for individual students, and roster verification systems).